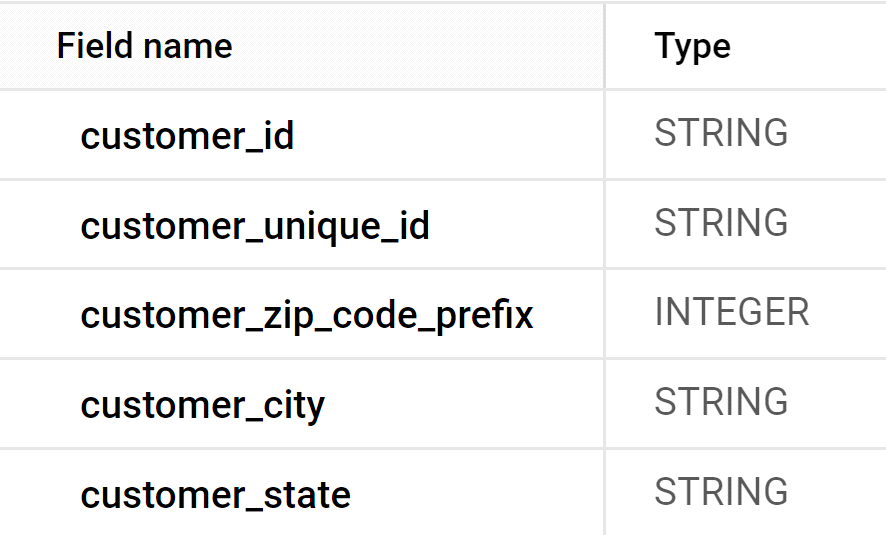
Business Case Project

**1. 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.

DESCRIBE

From business\_data.customers;



**B**. Get the time range between which the orders were placed.

Select

min (order\_purchase\_timestamp) as first\_order,

max(order\_purchase\_timestamp) as last\_order

From business\_data.orders;



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

with demo as (

select \* from business\_data.orders as o

join business\_data.customers as c

on o.customer\_id = c.customer\_id)

select

count(distinct customer\_city) as num\_unique\_cities,

count(distinct customer\_state) as num\_unique\_states

from demo

where order\_purchase\_timestamp between "2016-09-04" and "2018-10-17”;



II.In-depth Exploration:

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

Select

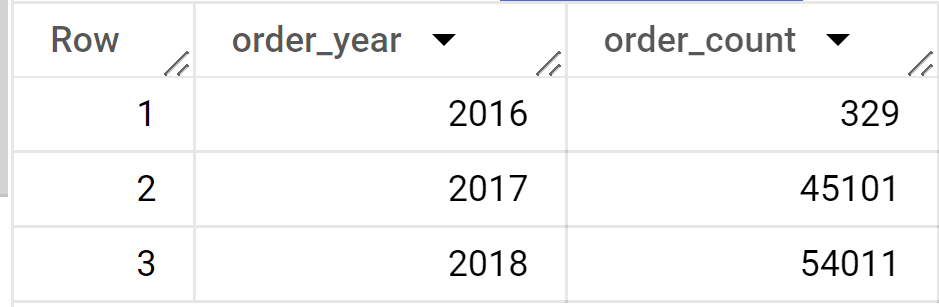
Extract (year from or der\_purchase\_timestamp) as order\_year,

Count(\*) as order\_count

From business\_data.orders

Group by (year from order\_purchase\_timestamp)

Order by order\_year;



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

Select

extract(year from order\_purchase\_timestamp) as year,

extract(month from order\_purchase\_timestamp) as month,

count(\*) as order\_count

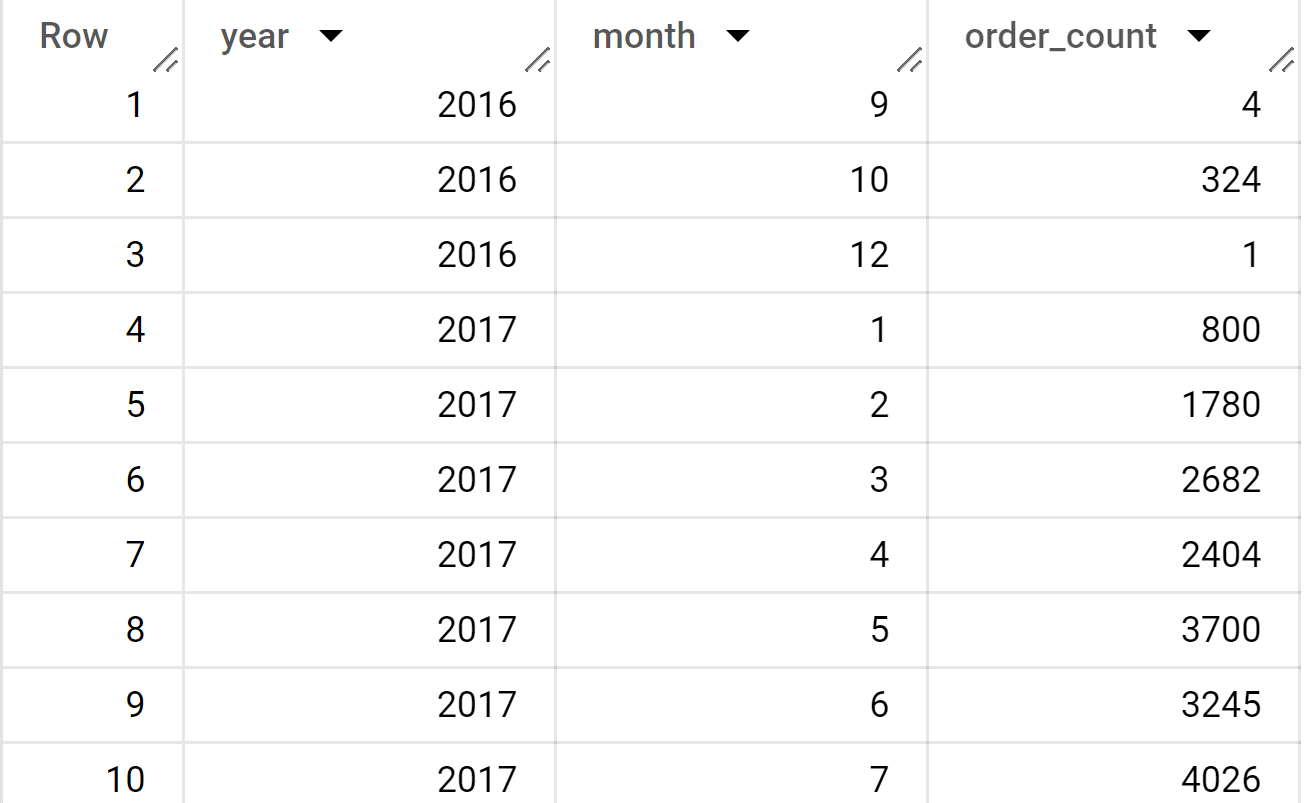
from business\_data.orders

Group by

extract(year from order\_purchase\_timestamp),

extract(month from order\_purchase\_timestamp)

order by year,month;



**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.

Select c.customer\_city,

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 "morning"

when extract(hour from order\_purchase\_timestamp) between 13 and 18 then "afternoon"

else "night"

end as order\_time\_category,

count(\*) as order\_count

from business\_data.orders as o

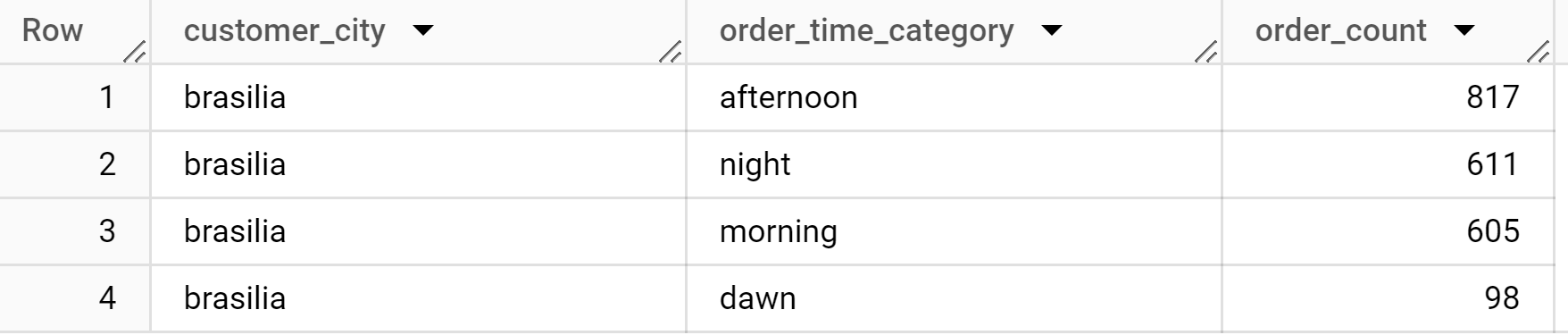
 join business\_data.customers as c

 on o.customer\_id = c.customer\_id

where c.customer\_city = "brasilia"

group by order\_time\_category,c.customer\_city

order by order\_count desc;



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

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

With demo as (

Select \*

from business\_data.customers as c

Right join business\_data.orders as o

On c.customer\_id=o.customer\_id)

Select extract ( month from order\_purchase\_timestamp) as month,

customer\_state,

count (\*)as num\_orders

From demo

Group by month,customer\_state

Order by month,customer\_state;



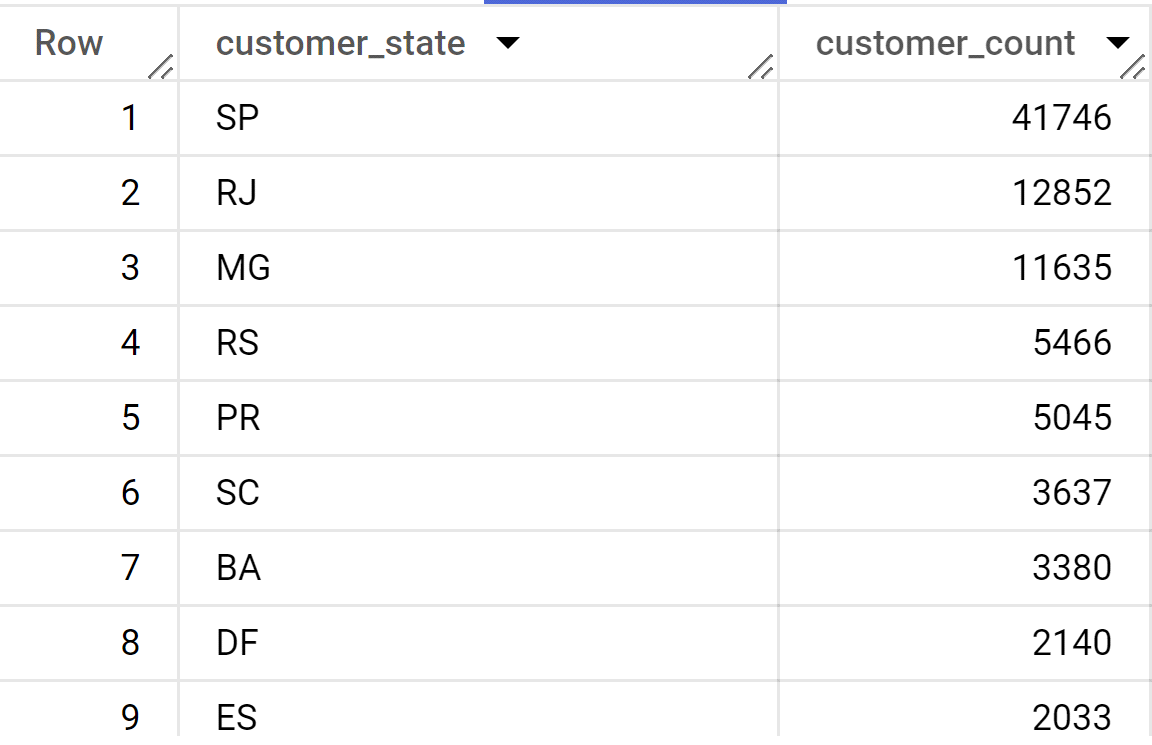
**B**. How are the customers distributed across all the states?

Select Customer\_state,count (\*)as customer\_count

From business\_data.customers

Group by customer\_state

Order by customer\_count desc;



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

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

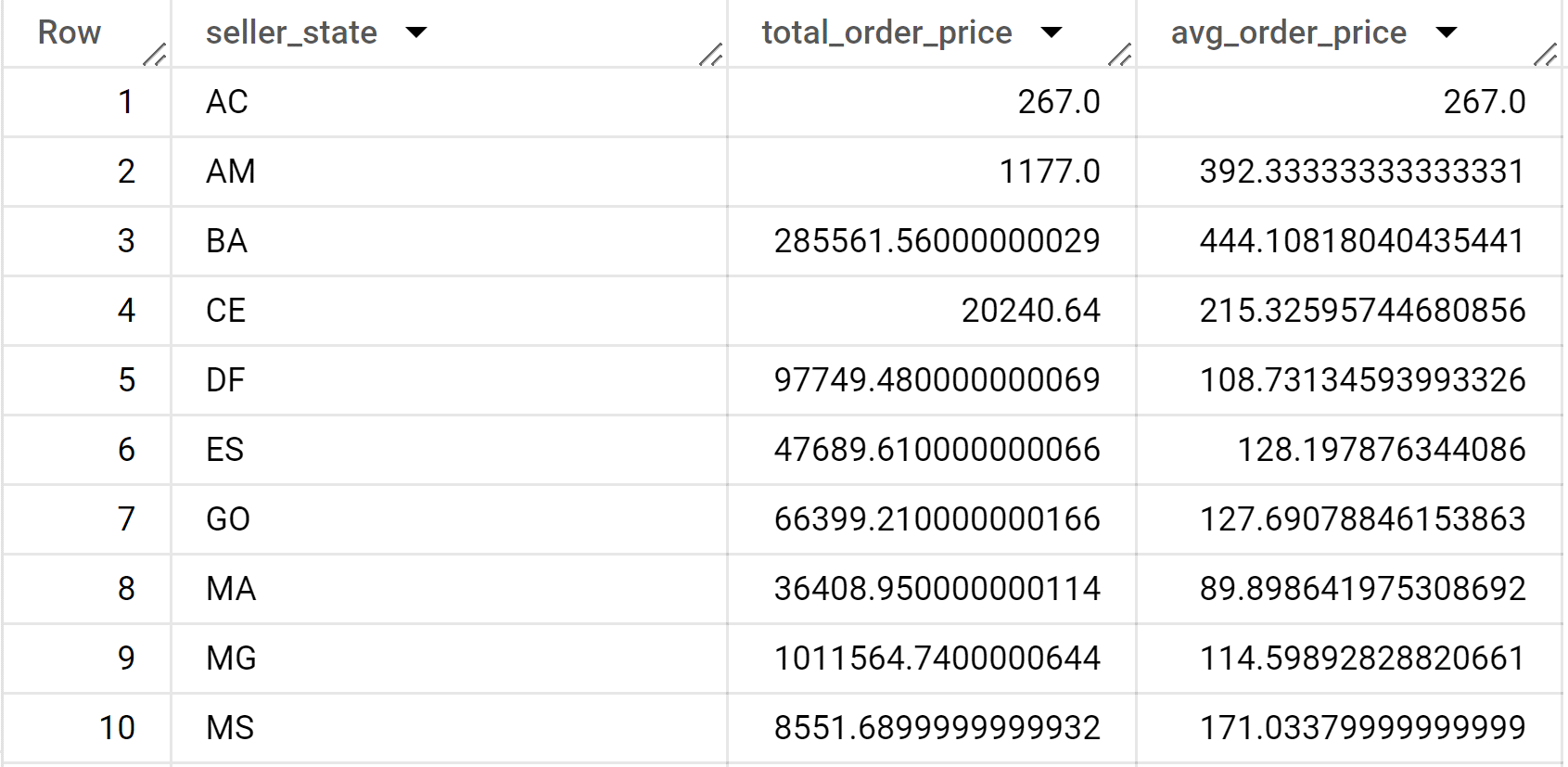
Select s.seller\_state, sum(price) AS total\_order\_price,

avg (price)as avg\_order\_price from business\_data.sellers as s

Join business\_data.order\_items as o

On s.seller\_id = o.seller\_id

Group by s.seller\_state;



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

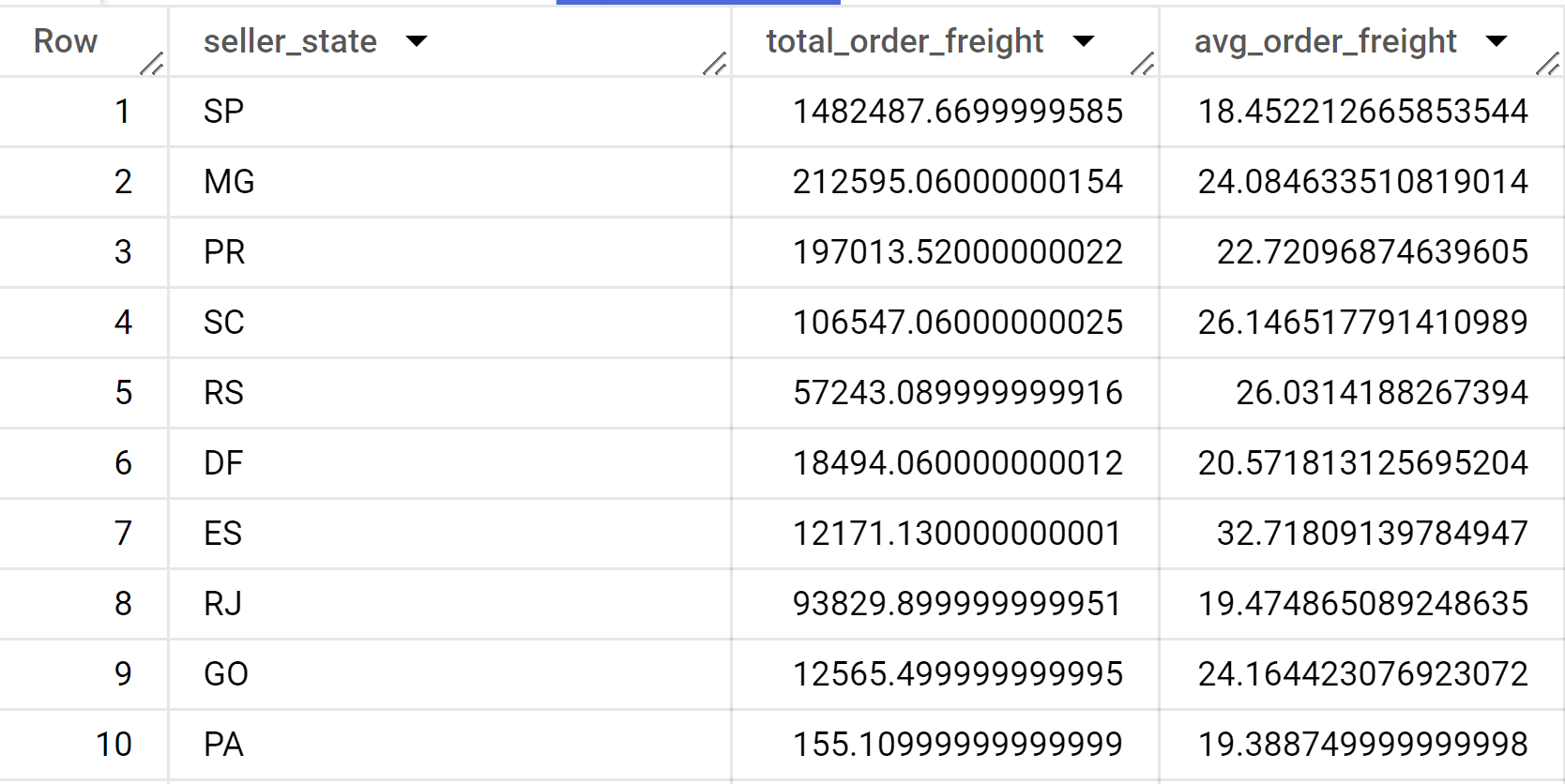
Select s.seller\_state,sum(freight\_value) as total\_order\_freight,

avg(freight\_value) as avg\_order\_freight from business\_data.order\_items as o

join business\_data.sellers as s

on o.seller\_id = s.seller\_id

group by s.seller\_state;



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

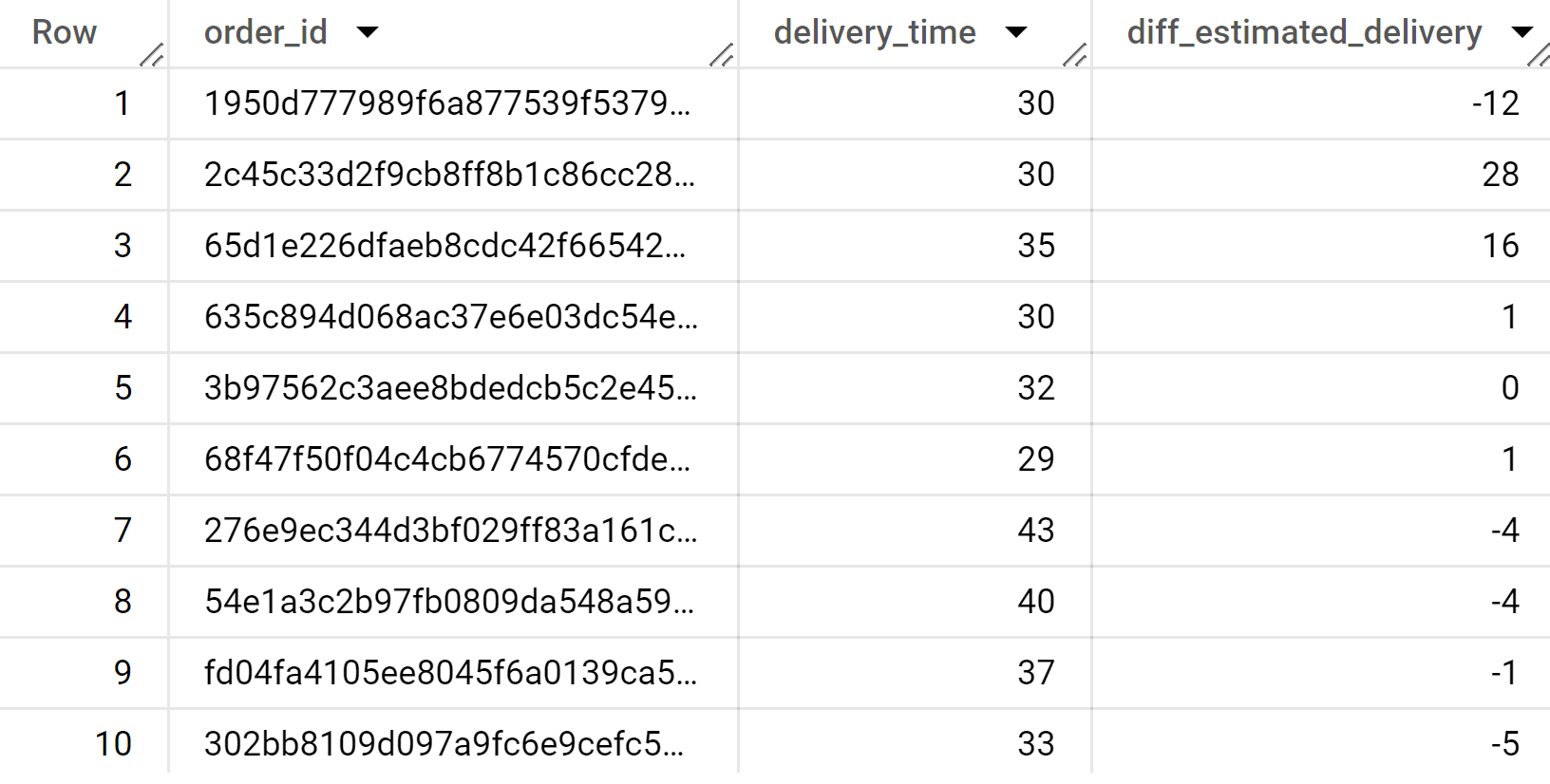
**A**.Find the no. of days taken to deliver each order from the order’s purchase date as delivery time. Also, calculate the difference (in days(between the estimated & actual delivery date of an order. Do this in a single query.

Select order\_id,

DATE\_DIFF(order\_delivered\_customer\_date,order\_purchase\_timestamp,DAY)as delivery\_time,

DATE\_DIFF(order\_estimated\_delivery\_date,order\_delivered\_customer\_date,DAY)as diff\_estimated\_delivery

from business\_data.orders;



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

**TOP 5 state highest average freight value**

Select s.seller\_state,avg (freight\_value) as avg\_freight\_value

from business\_data.order\_items as o

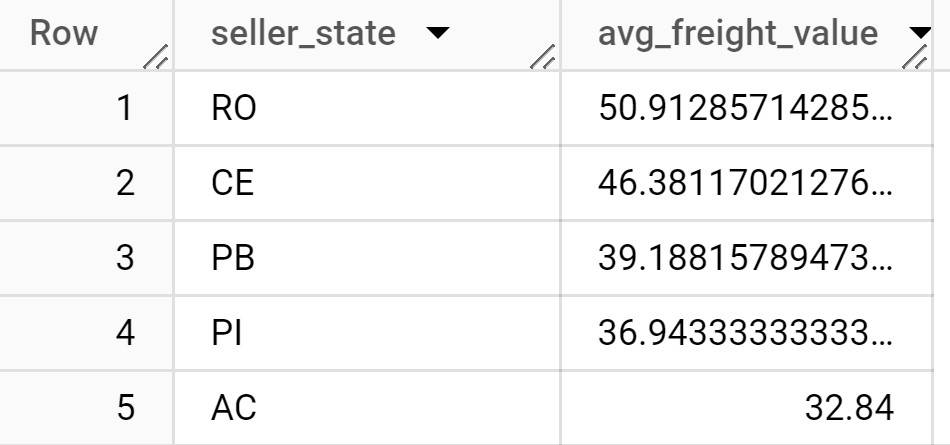
full join business\_data.sellers as s

 on o.seller\_id = s.seller\_id

 group by s.seller\_state

 order by avg\_freight\_value desc

 limit 5;



**TOP 5 state lowest average freigh value**

select s.seller\_state,avg(freight\_value) as avg\_freight\_value

from business\_data.order\_items as o

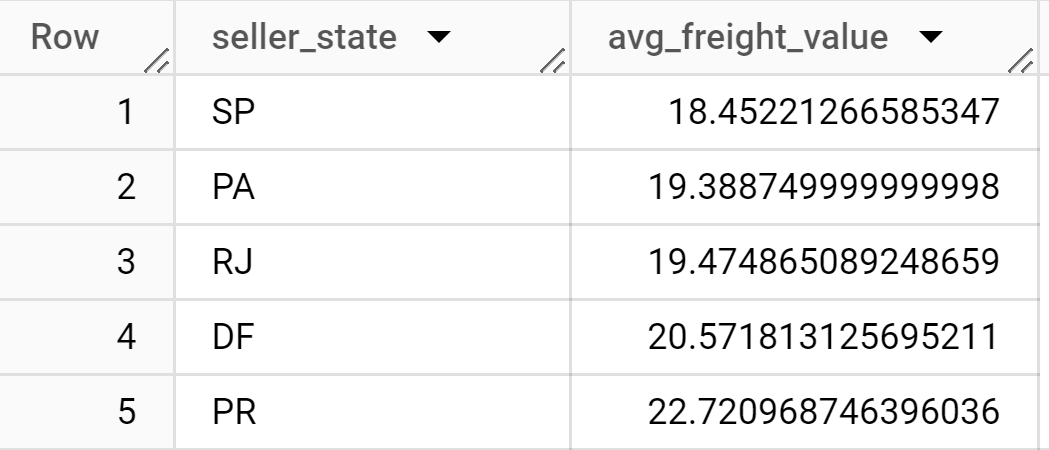
full join business\_data.sellers as s

on o.seller\_id = s.seller\_id

group by s.seller\_state

order by avg\_freight\_value asc

limit 5;



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

Top 5 state highest average delivery time

with demo as (

select \* from business\_data.orders as o

join business\_data.customers as c

on o.customer\_id = c.customer\_id)

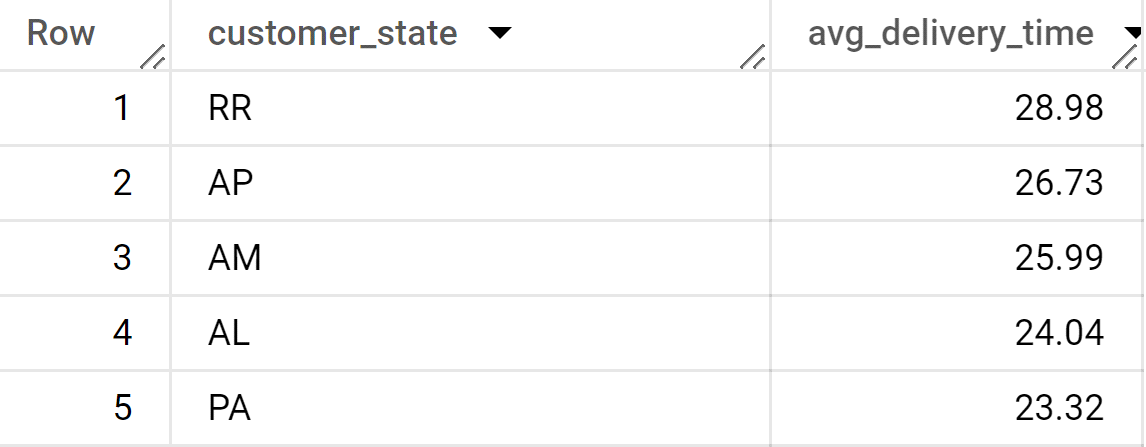
select customer\_state,round(avg(DATE\_DIFF(order\_delivered\_customer\_date,order\_purchase\_timestamp,HOUR)),2) as avg\_delivery\_time

from demo

group by customer\_state

order by avg\_delivery\_time asc

limit 5;



Top 5 state lowest average delivery time

with demo as (

select \* from business\_data.orders as o

join business\_data.customers as c

on o.customer\_id = c.customer\_id

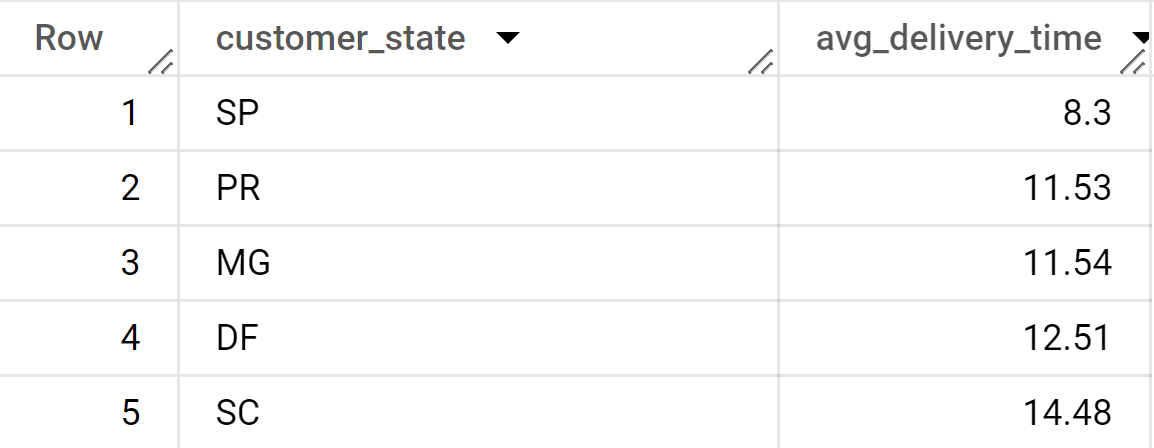
)

select customer\_state,round(avg(DATE\_DIFF(order\_delivered\_customer\_date,order\_purchase\_timestamp,DAY)),2) as avg\_delivery\_time

from demo

group by customer\_state

order by avg\_delivery\_time desc limit 5;



**D**. 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.

with demo as

(select \* from business\_data.orders as o

join business\_data.customers as c

on o.customer\_id = c.customer\_id)

select customer\_state,

date\_diff(order\_delivered\_customer\_date,order\_purchase\_timestamp,day)-

date\_diff(order\_estimated\_delivery\_date,order\_purchase\_timestamp, day)

as delivery\_speed

from demo

where order\_status="delivered"

group by customer\_state,delivery\_speed

order by delivery\_speed desc

limit 5;



VI. Analysis based on the payments:

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

select

extract(year from o.order\_purchase\_timestamp) as year,

extract(month from o.order\_purchase\_timestamp) as month,

p.payment\_type,

count(\*) as num\_orders

from business\_data.orders as o

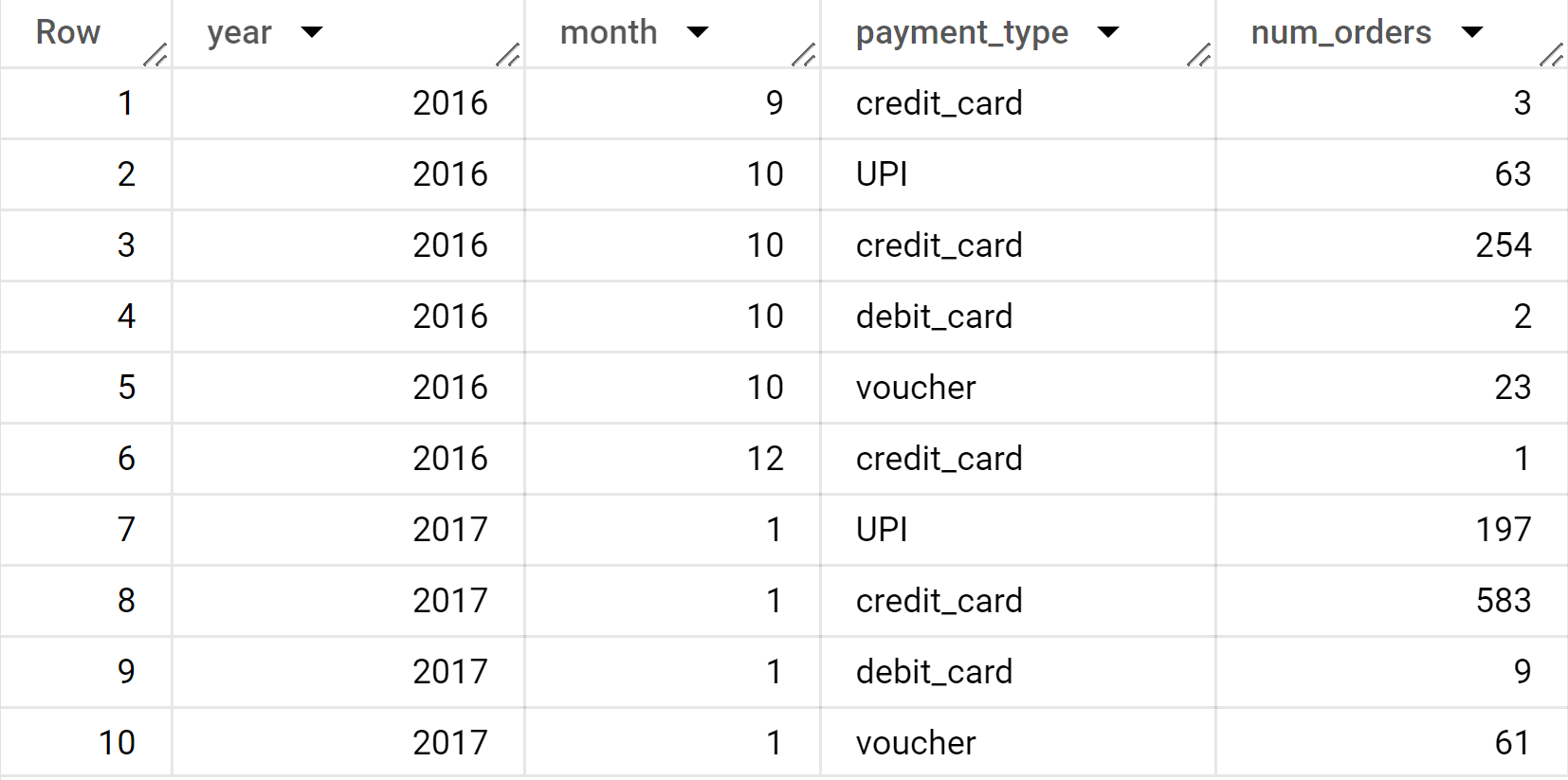
join business\_data.payments as p

on o.order\_id = p.order\_id

group by year,month,p.payment\_type

order by year,month,p.payment\_type

limit 10;



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

Select count(distinct order\_id) as num\_orders

from business\_data.payments

where payment\_installments =1;

