CASE STUDY OF TARGET

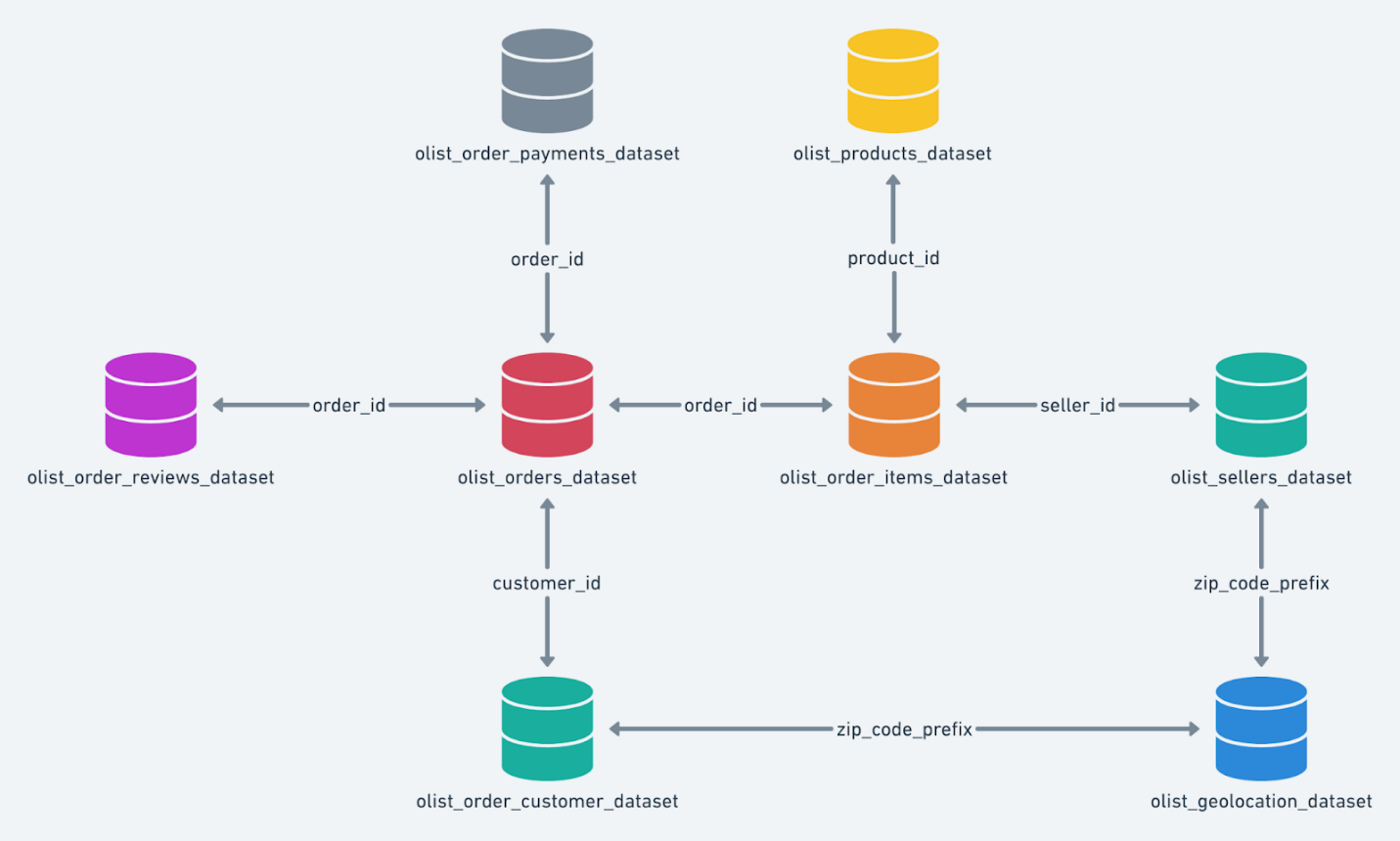
Target is a globally renowned brand and a prominent retailer in the United States. Target makes itself a preferred shopping destination by offering outstanding value, inspiration, innovation and an exceptional guest experience that no other retailer can deliver.

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:**

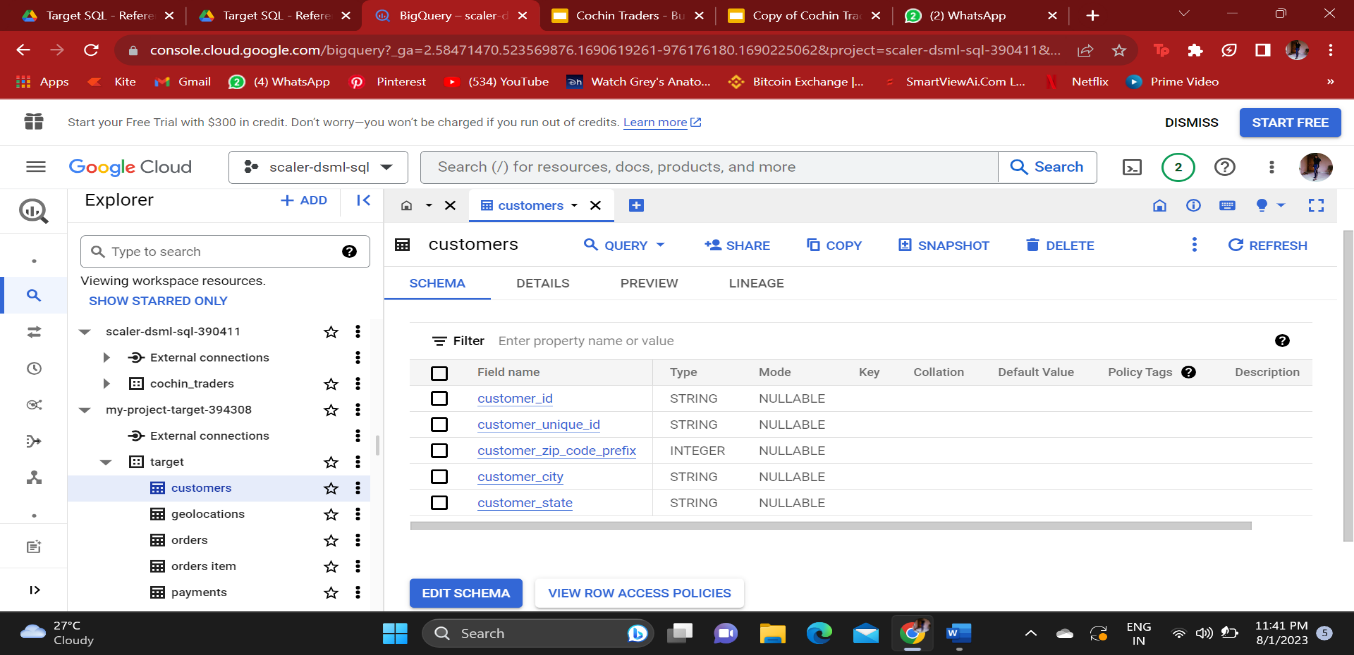
Target, a prominent retail chain, strives to optimize its operations, enhance customer satisfaction, and maximize profitability. To achieve these goals, it is essential to gain actionable insights from the vast amount of data generated across various departments and touchpoints. This study aims to conduct a comprehensive dataset analysis for Target stores in the UK. The analysis will focus on understanding customer behavior, inventory management, sales patterns, and market trends specific to the UK retail market. By delving into this data, the research intends to identify strategic opportunities, operational inefficiencies, and areas for improvement within Target's UK stores.



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

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 Cities & States of customers who ordered during the given period.
2. 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.

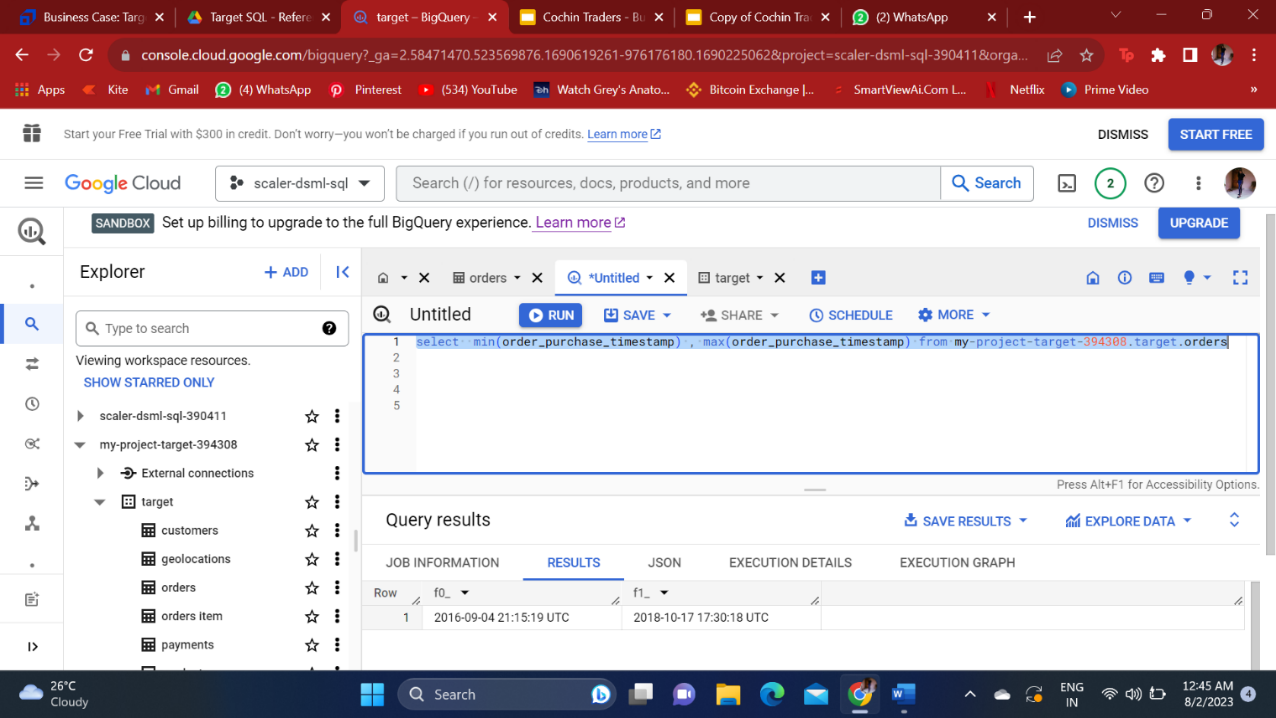


1. Get the time range between which the orders were placed.

Answer -

Query : SELECT min(order\_purchase\_timestamp), max(order\_purchase\_timestamp)

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



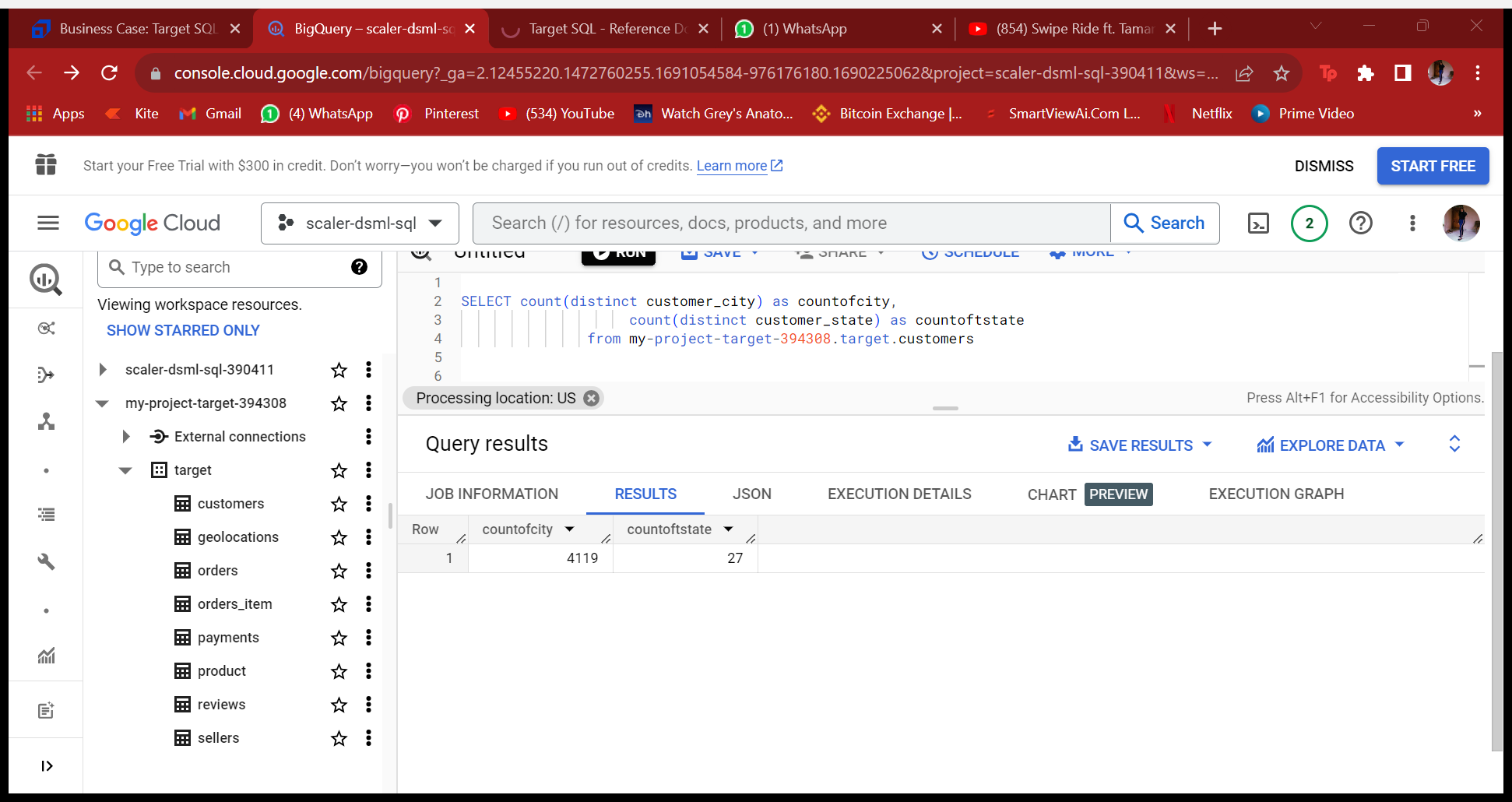
1. Count the Cities & States of customers who ordered during the given period.

Answer :

Query: SELECT count(distinct customer\_city) as countofcity,

                    count(distinct customer\_state) as countoftstate

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



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

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

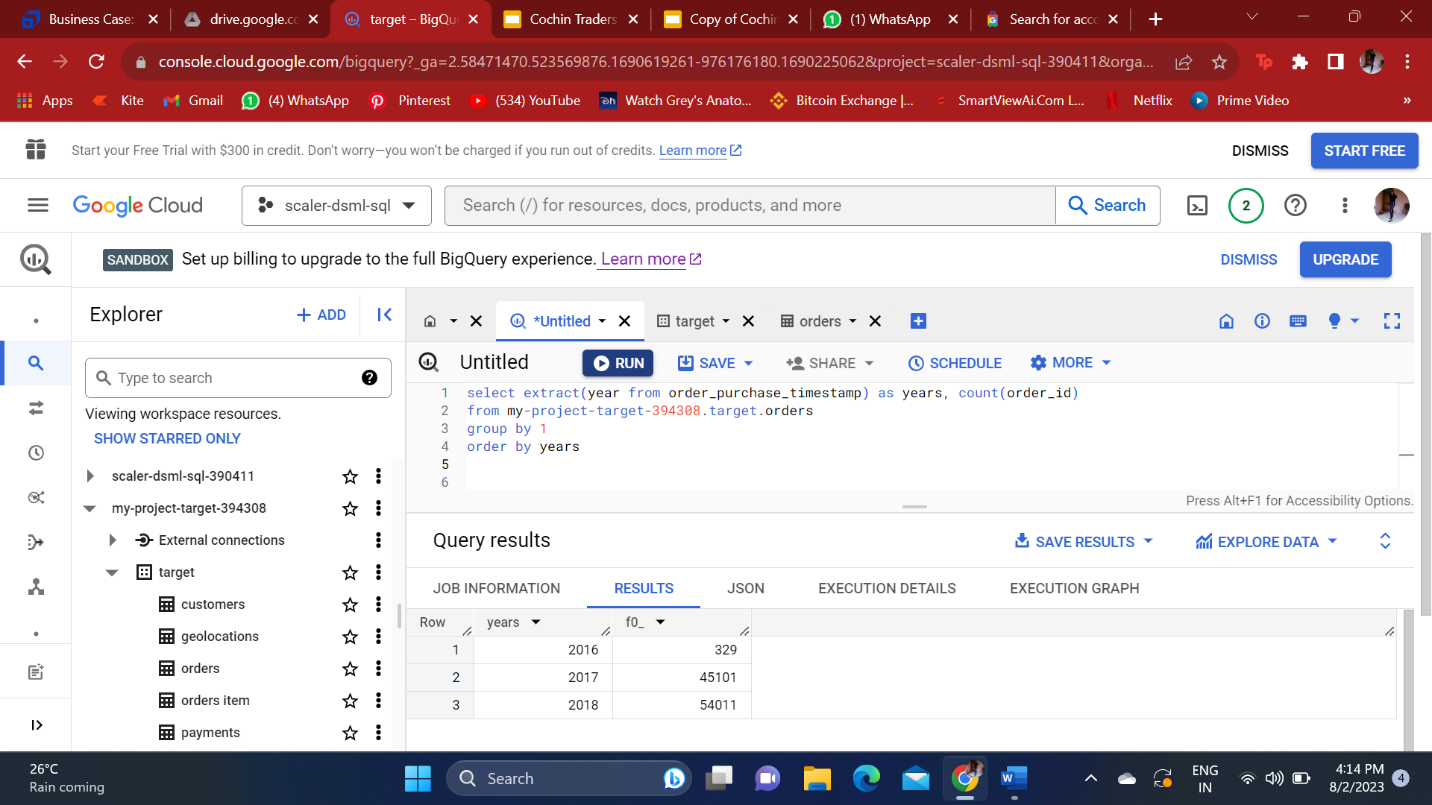
Answer-

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



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.

* 1. 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

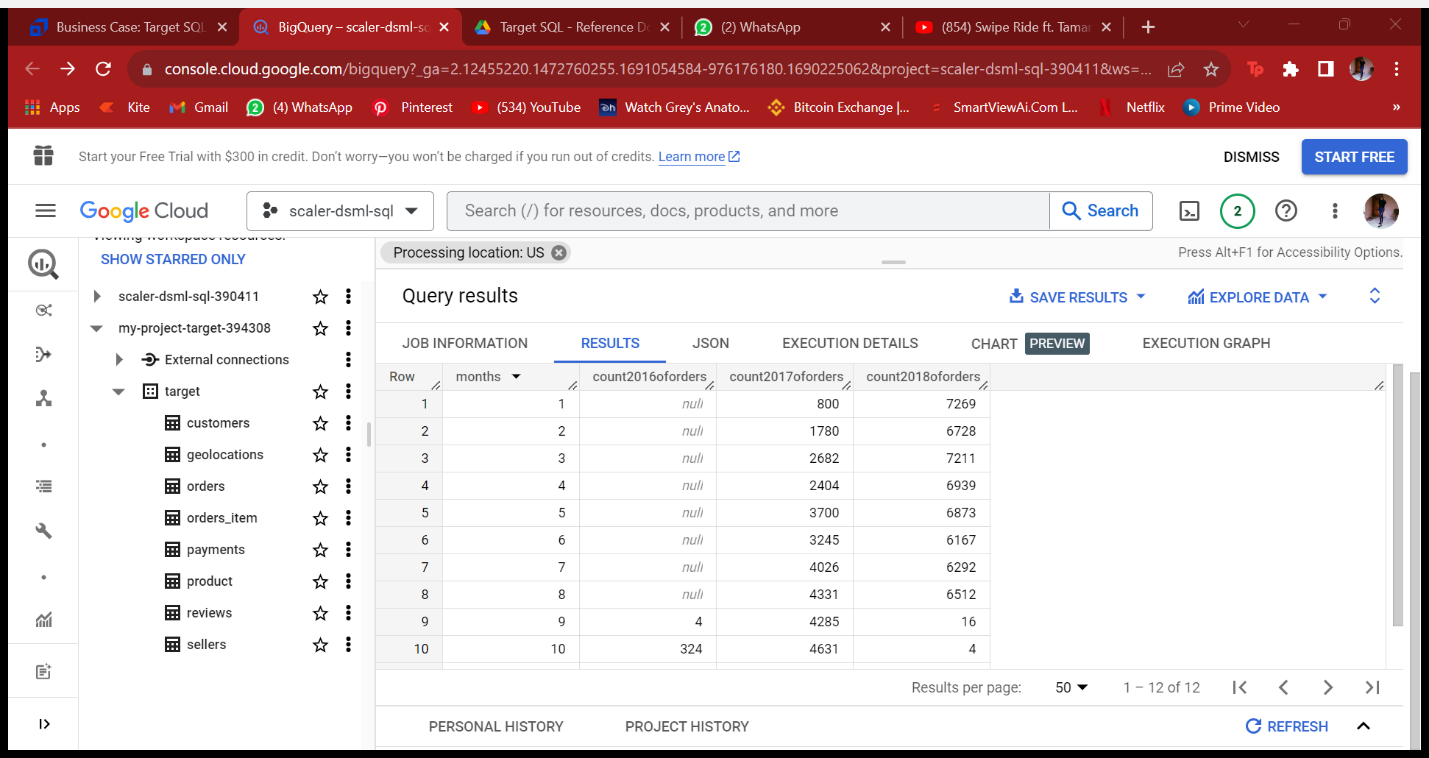
                group by months , years

                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



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'

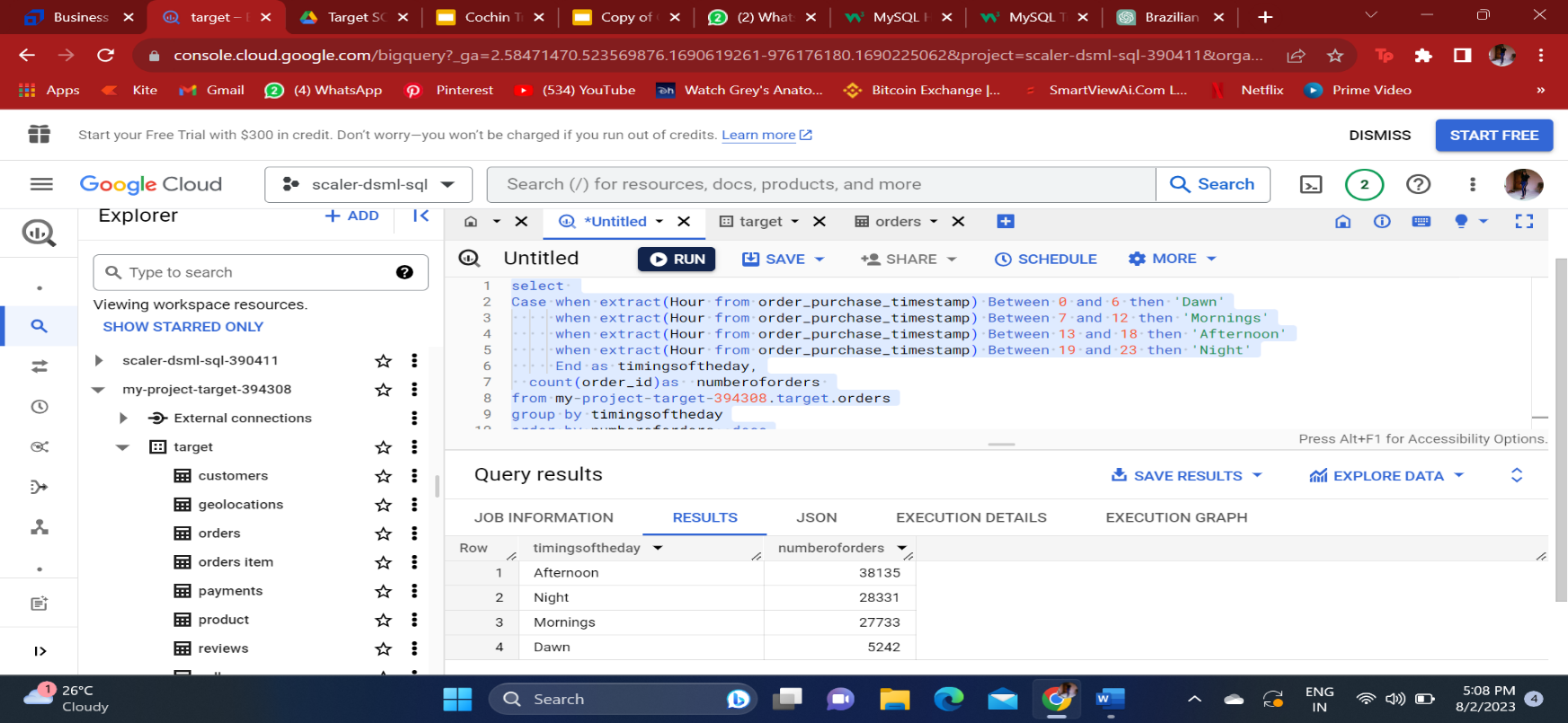
when extract (Hour from order\_purchase\_timestamp) Between 19 and 23 then 'Night'

End as timingsoftheday, count (order\_id)as numberoforders

from my-project-target-394308.target.orders

group by timingsoftheday

order by numberoforders desc



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

1. **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?
2. 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,

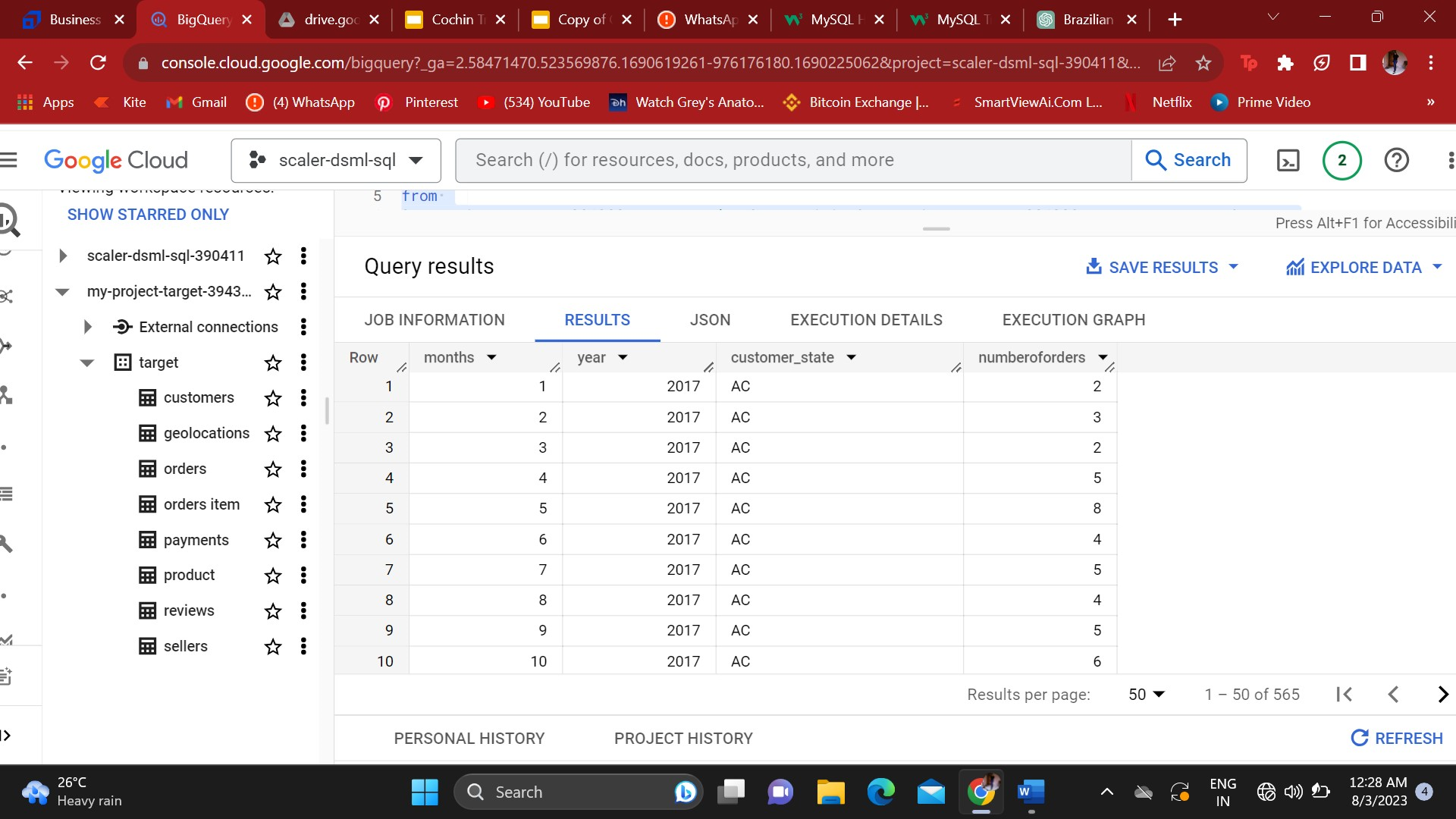
  customer\_state,

  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



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

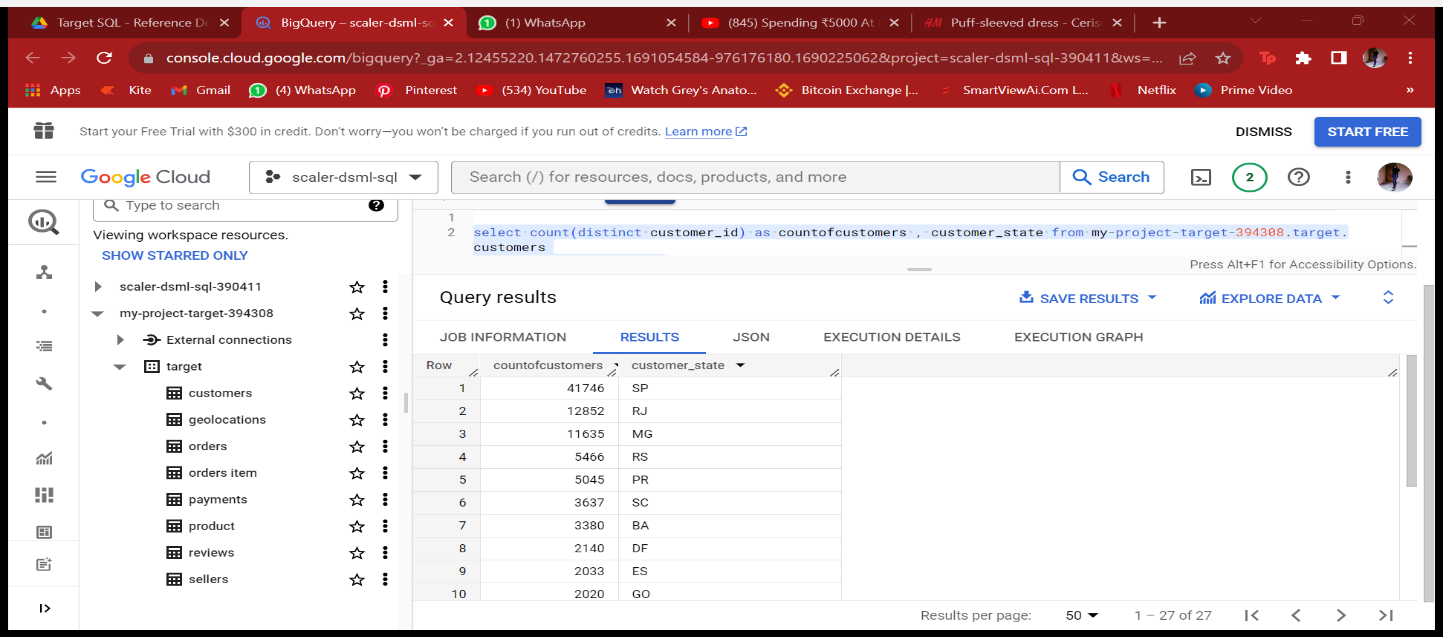
select count(distinct customer\_id) as countofcustomers ,

customer\_state

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

group by customer\_state

order by countofcustomers desc



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

1. **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. 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

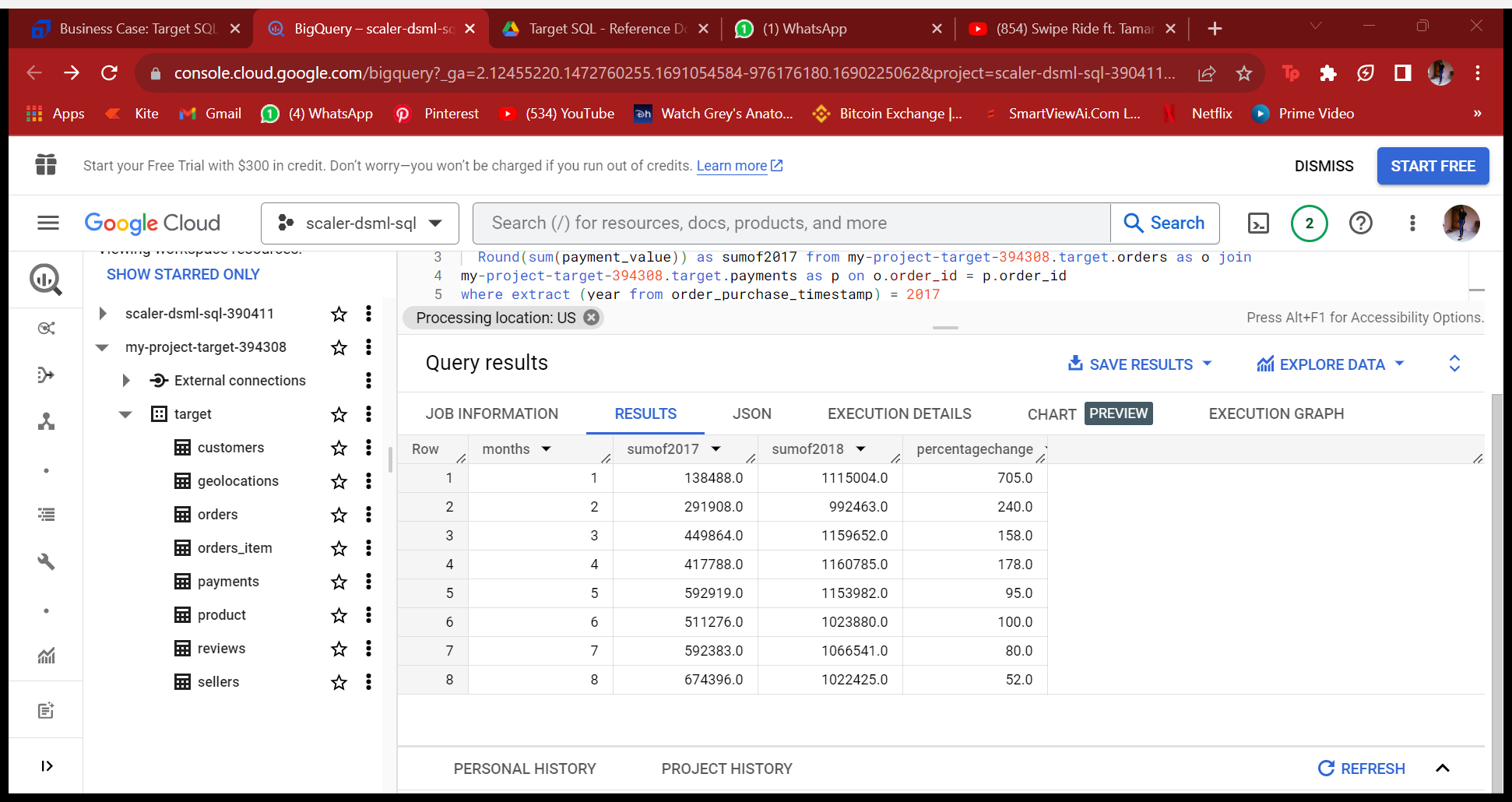
order by months)

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



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.

* + 1. 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

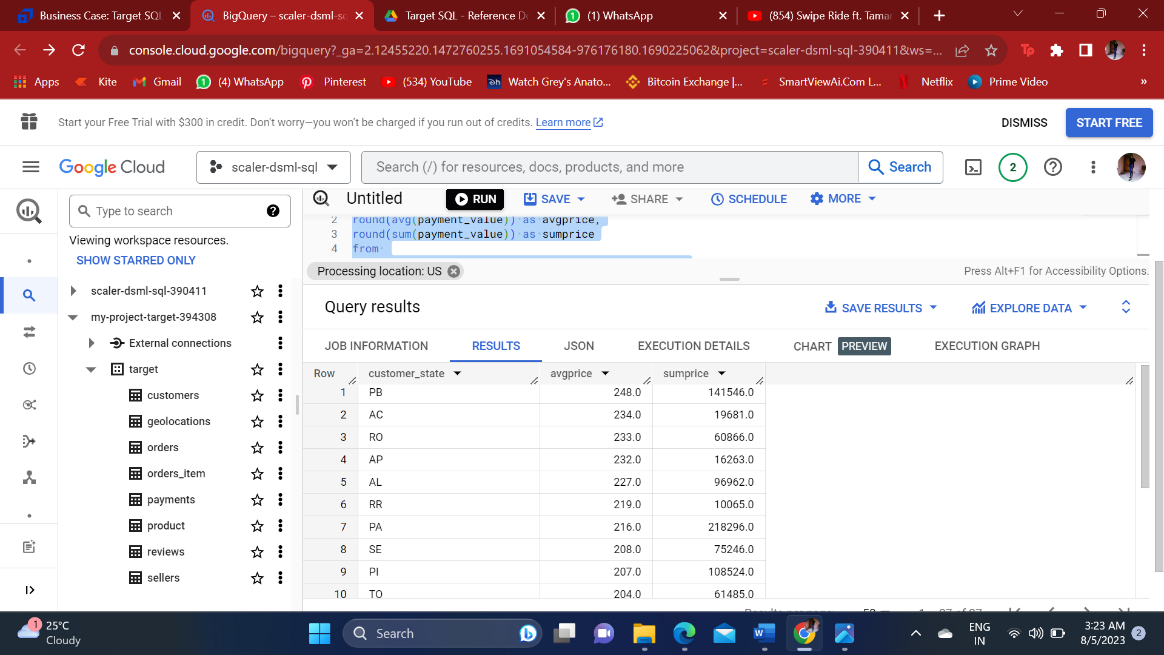
on c.customer\_id = o.customer\_id

join my-project-target-394308.target.payments as p

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

group by 1

order by avgprice desc



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

* + 1. 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

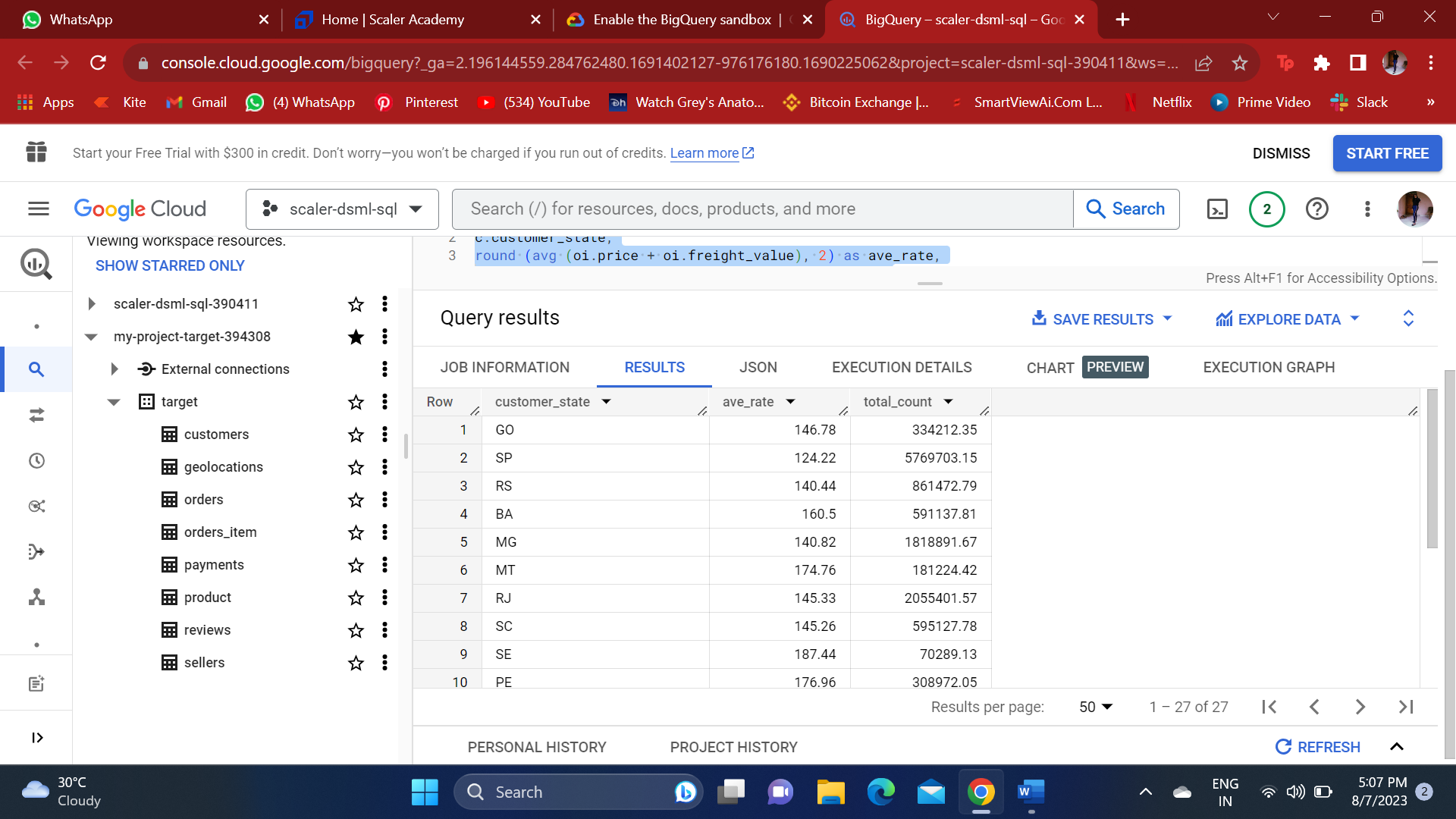
join

 my-project-target-394308.target.orders\_item as oi

on o.order\_id = oi.order\_id

where o.order\_status = 'delivered'

group by 1



1. **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



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

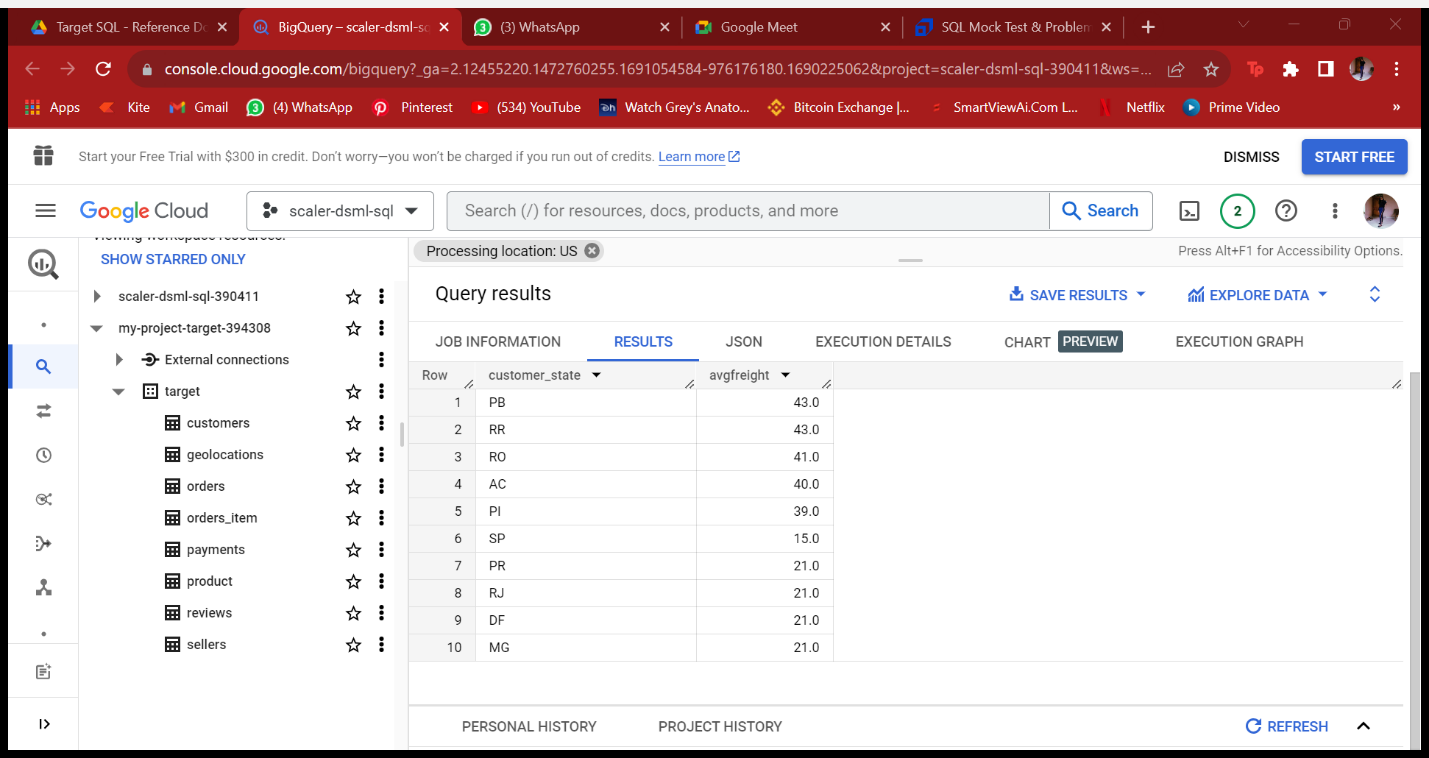
order by avgfreight asc

limit 5 )

select \* from highest

union all

select \* from lowest



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

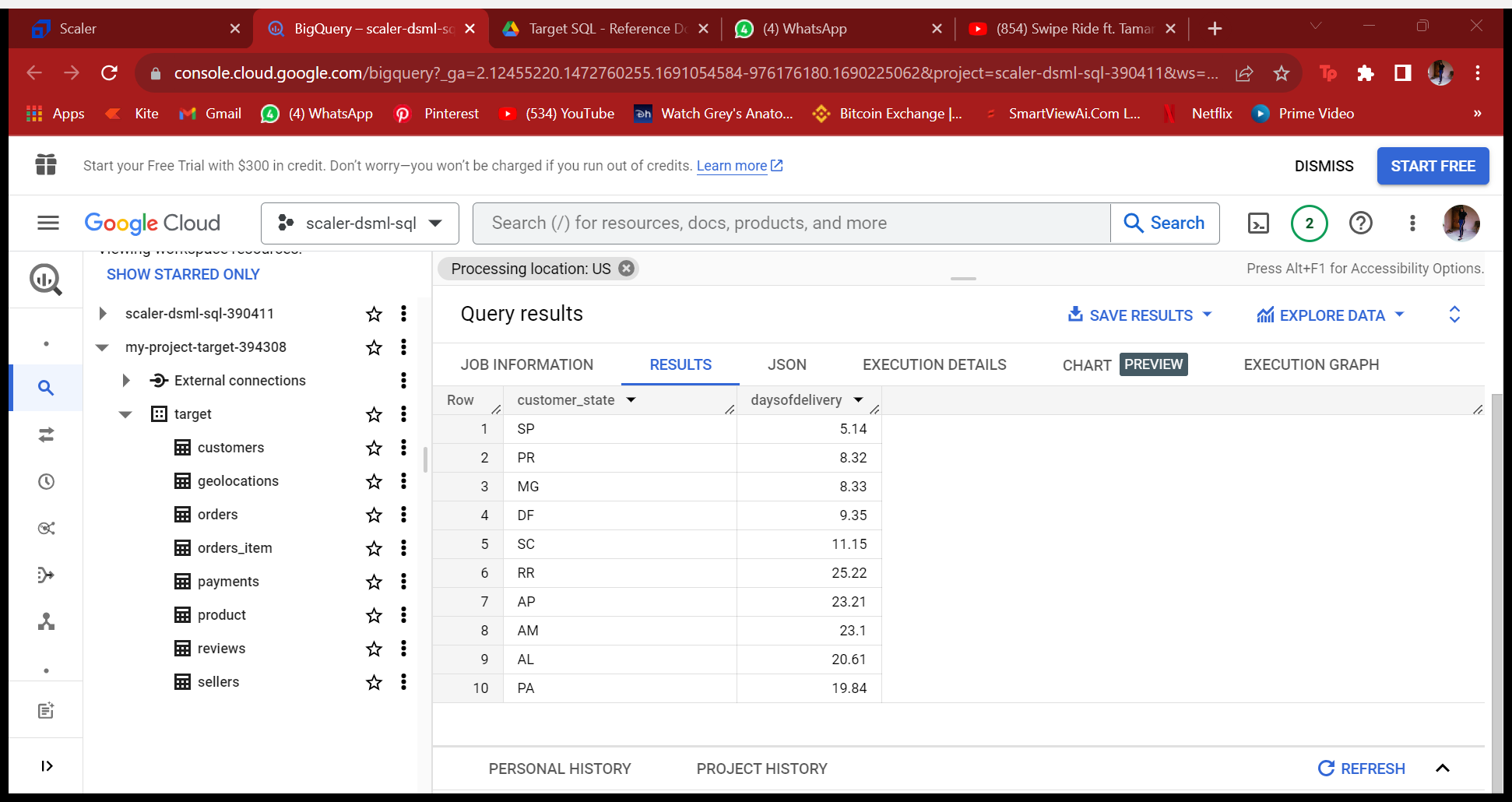
order by daysofdelivery desc

limit 5 )

select \* from highest

union all

select \* from lowest



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

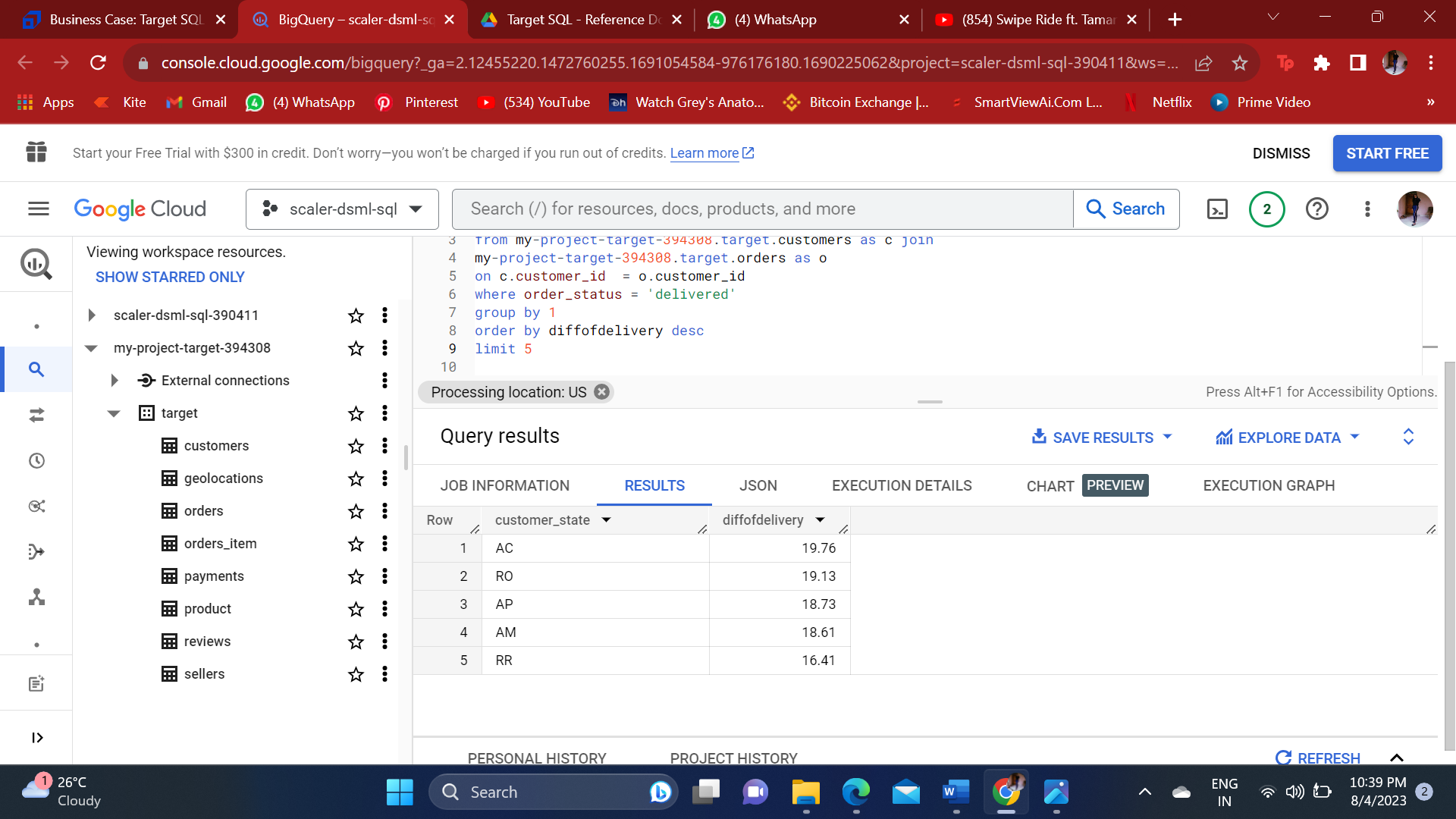
on c.customer\_id  = o.customer\_id

where order\_status = 'delivered'

group by 1

order by diffofdelivery desc

limit 5



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:

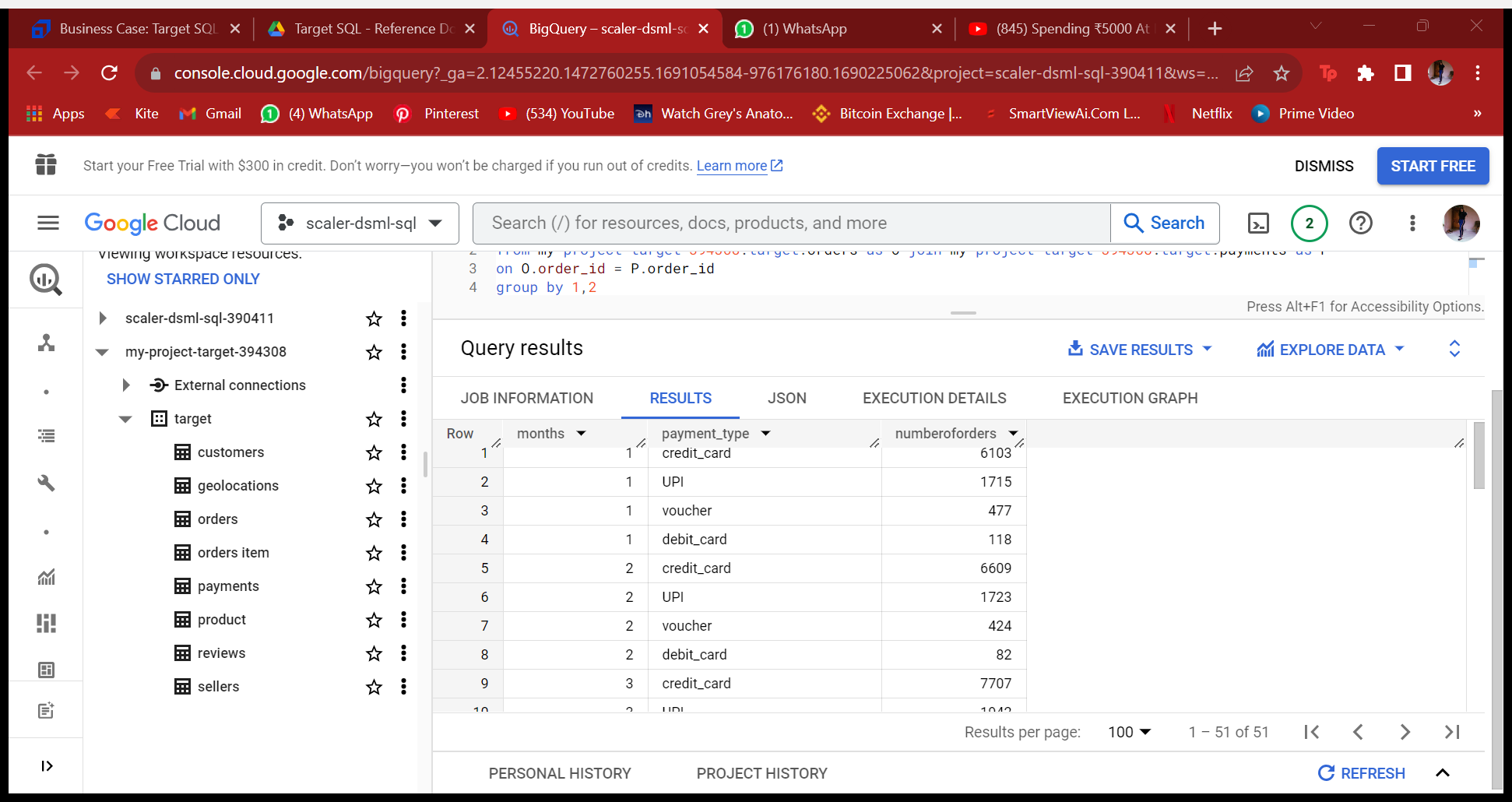
select extract (month from order\_purchase\_timestamp) as months, payment\_type, count(O.order\_id) numberoforders

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



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

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

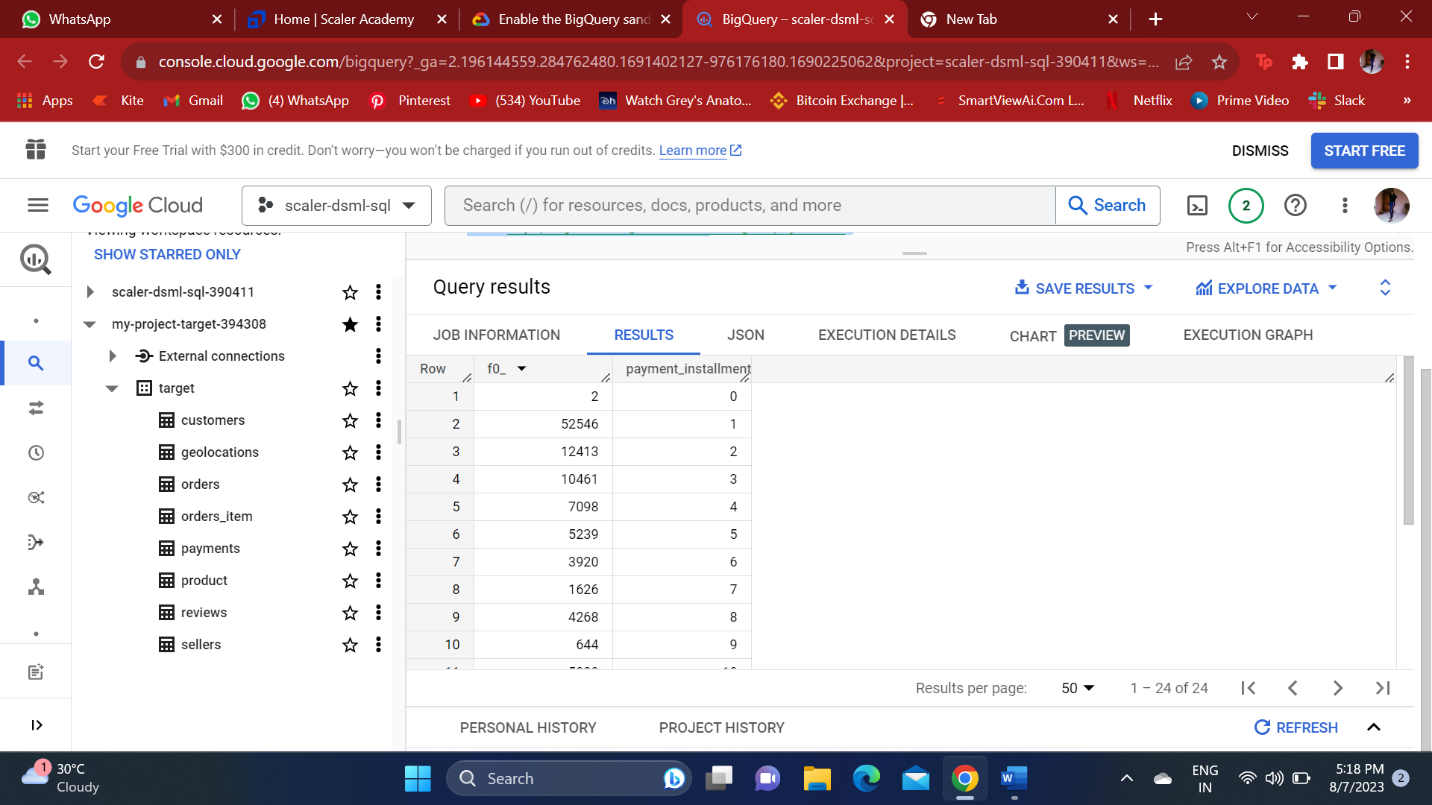
Answer-

Query :

select count(order\_id) , payment\_installments

from `my-project-target-394308.target.payments`

group by payment\_installments



**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