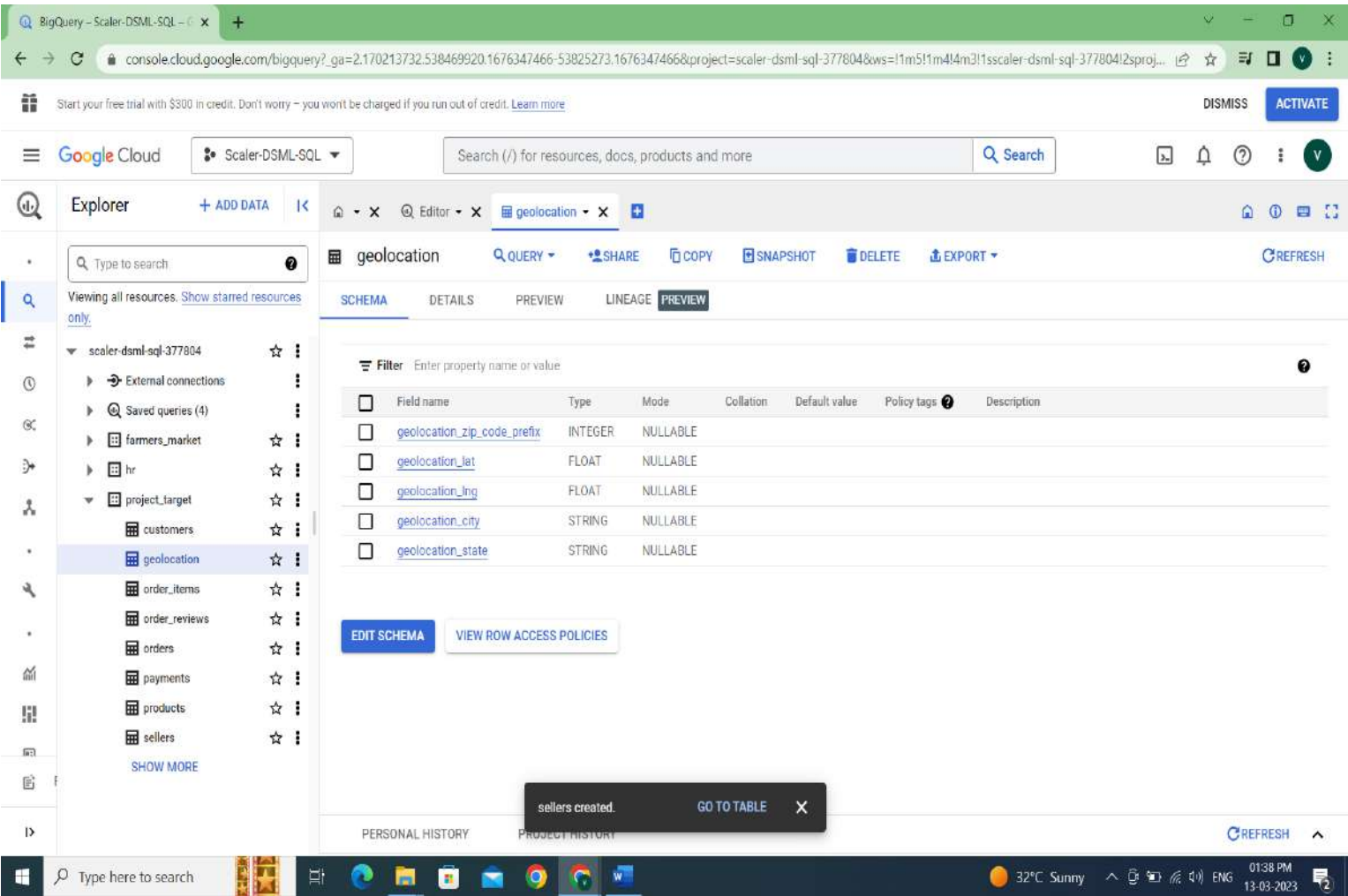
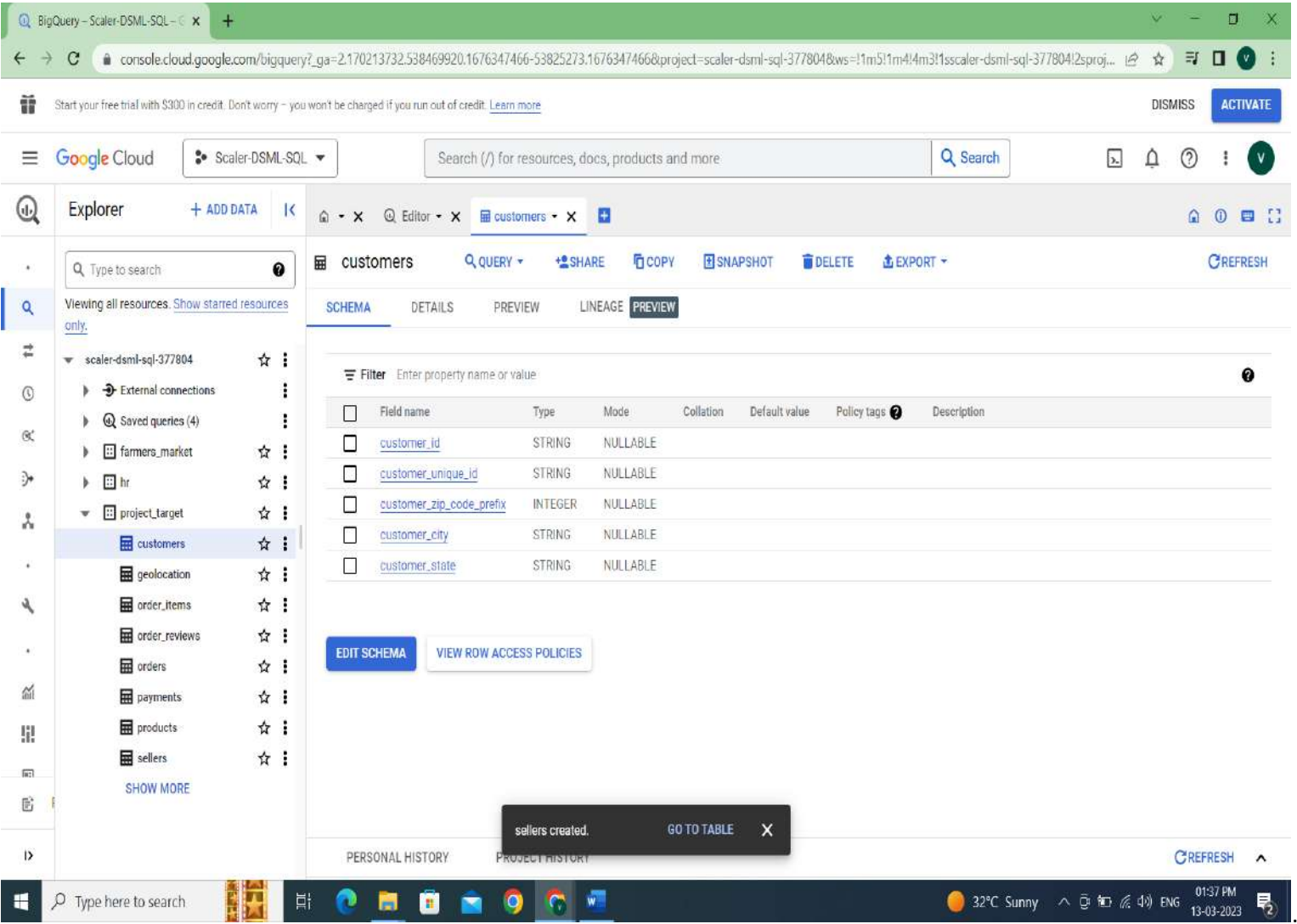


1. Import the dataset and do usual exploratory analysis steps like checking the structure & characteristics of the dataset

1. Data type of columns in a table?

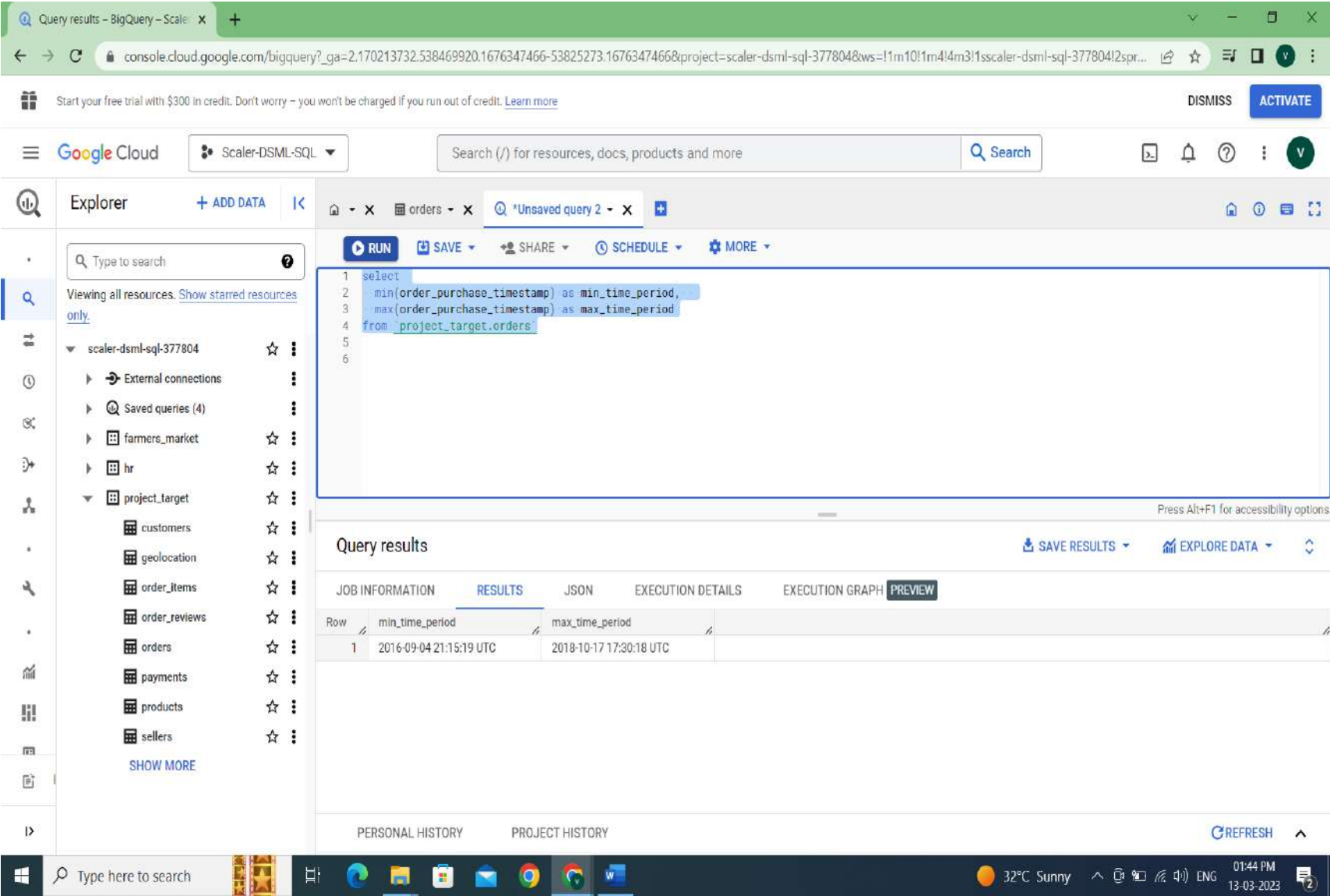
Ans: Data types of the column can be checked in big query checked by selecting the database and then selecting the table name. Post clicking the table, column name along with the data type will be displayed. Below is the SS of customer /geolocation table with data type. In the same way we can check other table’s as well.



2.Time period for which the data is given?

Ans:

```
select
min(order_purchase_timestamp) as min_time_period,
max(order_purchase_timestamp) as max_time_period
from `project_target.orders`
```



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

Ans:

```
select
distinct c.customer_state,
c.customer_city

from `project_target.customers` c
      join `project_target.orders` o
      on c.customer_id = o.customer_id
where order_purchase_timestamp
between '2016-09-04 21:15:19 UTC' and '2018-10-17 17:30:18 UTC'
order by 1,2
```

The screenshot displays the Google Cloud BigQuery console interface. At the top, there's a navigation bar with the Google Cloud logo, a search bar, and buttons for 'DISMISS' and 'ACTIVATE'. Below this, a 'SANDBOX' banner indicates a free trial with \$300 in credit. The main area is divided into three sections: Explorer, Editor, and Query results.

Explorer: Shows a list of resources under the project 'scaler-dsml-sql-377804'. The 'sql target project' is selected.

Editor: Contains a SQL query:

```

7  on c.customer_id = o.customer_id
8  where order_purchase_timestamp
9  between '2016-09-04 21:15:19 UTC' and '2018-10-17 17:30:18 UTC'
10 order by 1,2

```

Query results: Displays the results of the query in a table format. The table has two columns: 'customer_state' and 'customer_city'. The results are as follows:

Row	customer_state	customer_city
1	AC	brasileia
2	AC	cruzeiro do sul
3	AC	epitaciolandia
4	AC	manoel urbano
5	AC	porto acre
6	AC	rio branco
7	AC	senador guimard
8	AC	xapuri
9	AL	agua branca
10	AL	anadia

At the bottom, there are tabs for 'PERSONAL HISTORY' and 'PROJECT HISTORY', and a 'REFRESH' button.

Time period Analysis:

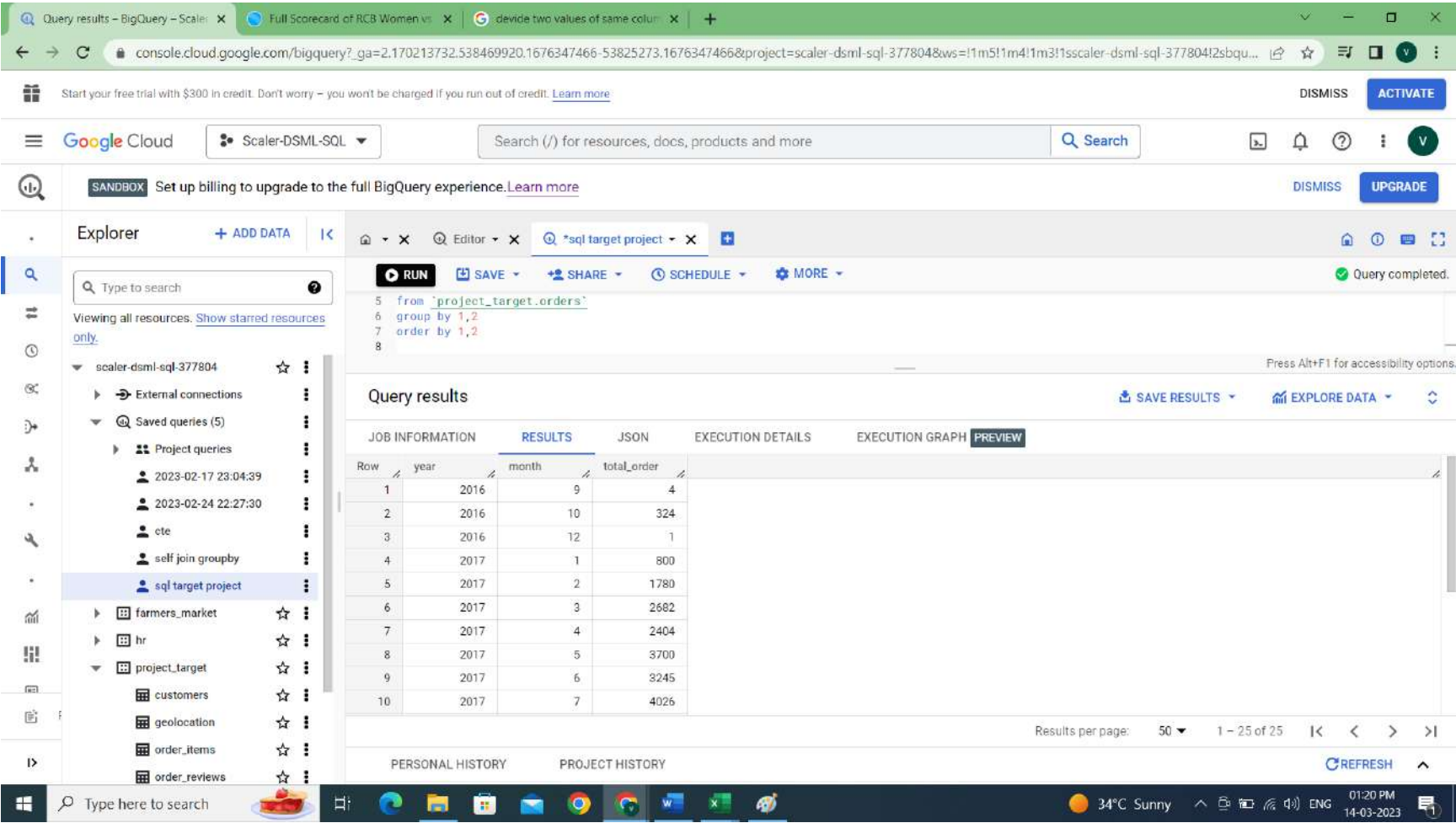
1. The data set has orders for **2 yrs. (Sep 2016-Oct 2018)** across Brazil from **27 states** which includes **4119 cities**.

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?

Ans:

```
select
  extract(year from order_purchase_timestamp) as year,
  extract(month from order_purchase_timestamp) as month,
  count(distinct order_id) total_order
from `project_target.orders`
group by 1,2
order by 1,2
```



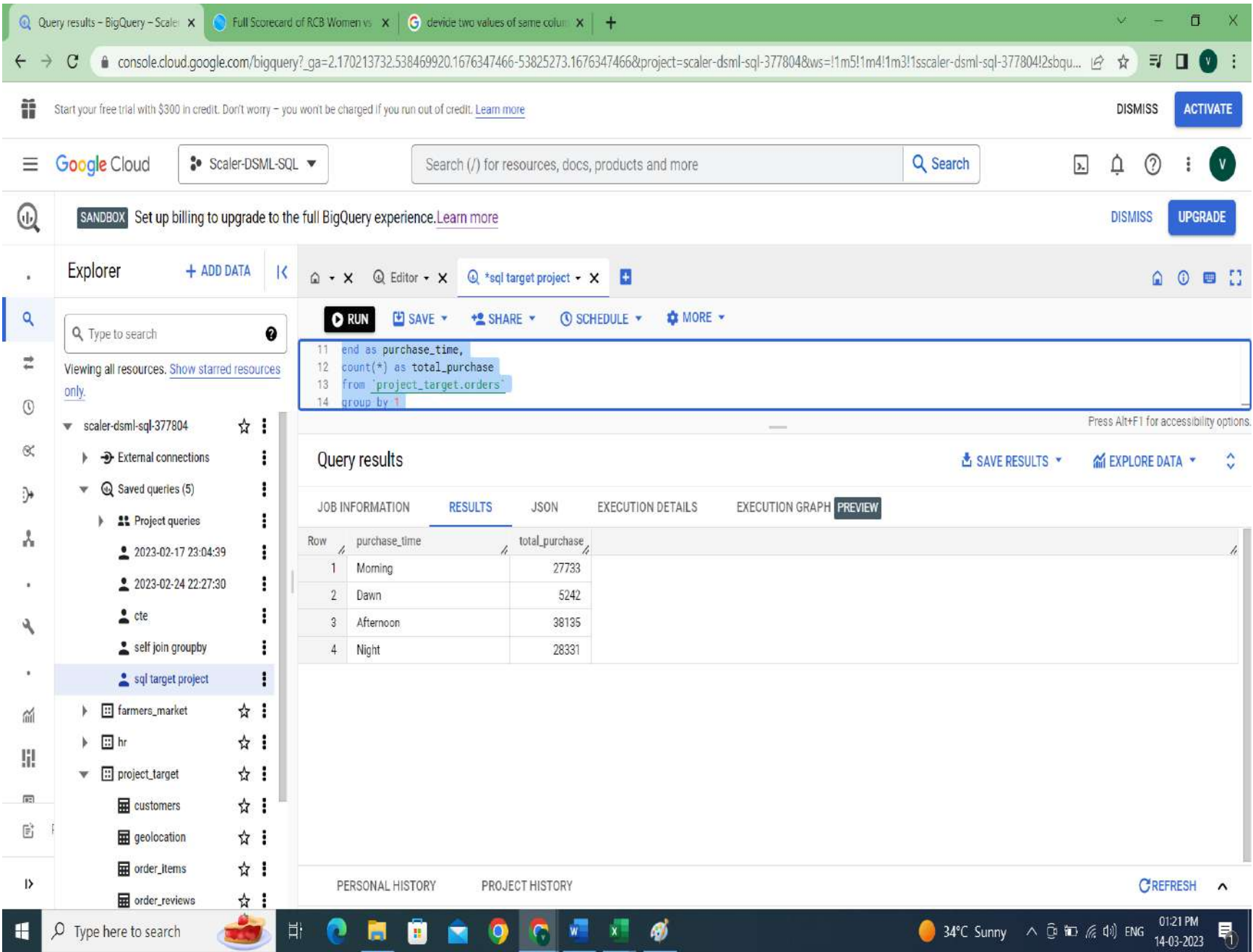
Trend Analysis:

- 1. We see the year and month wise data with order count, overall, there is a growing trend.
- 2. We can see that the sales count increase from month of **Jan till March**. This could be due to **carnival festival** held at Brazil every year around **last week of Feb till 1st Week of March**.

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

Ans:

```
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 'Morning'
  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 purchase_time,
count(*) as total_purchase
from `project_target.orders`
group by 1
```



Purchase time Analysis:

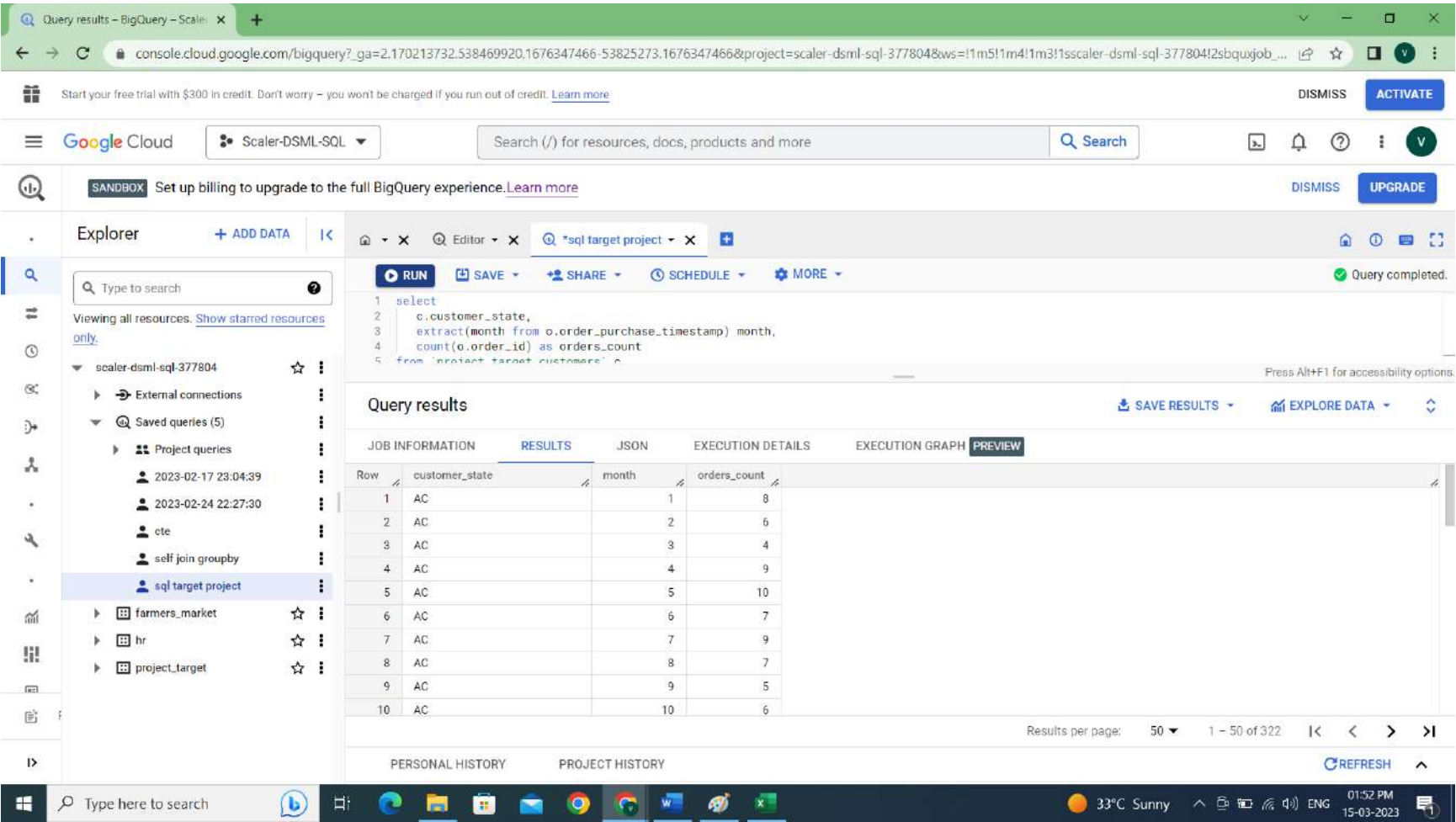
- a) We can see from the data that, majority of purchase happens on afternoon time(13:00 hrs. – 18:00 hrs.)
- b) Morning (07:00 hrs. – 12:00 hrs.) and night purchases (19:00-23:00) are almost similar.
- c) Least purchase happens at Dawn (0:00 hrs. – 06:00 hrs.)

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

1. Get month on month orders by states

Ans :

```
select
  c.customer_state,
  extract(month from o.order_purchase_timestamp) month,
  count(o.order_id) as orders_count
from `project_target.customers` c
  join `project_target.orders` o
  on c.customer_id = o.customer_id
group by 1,2
order by 1,2
```



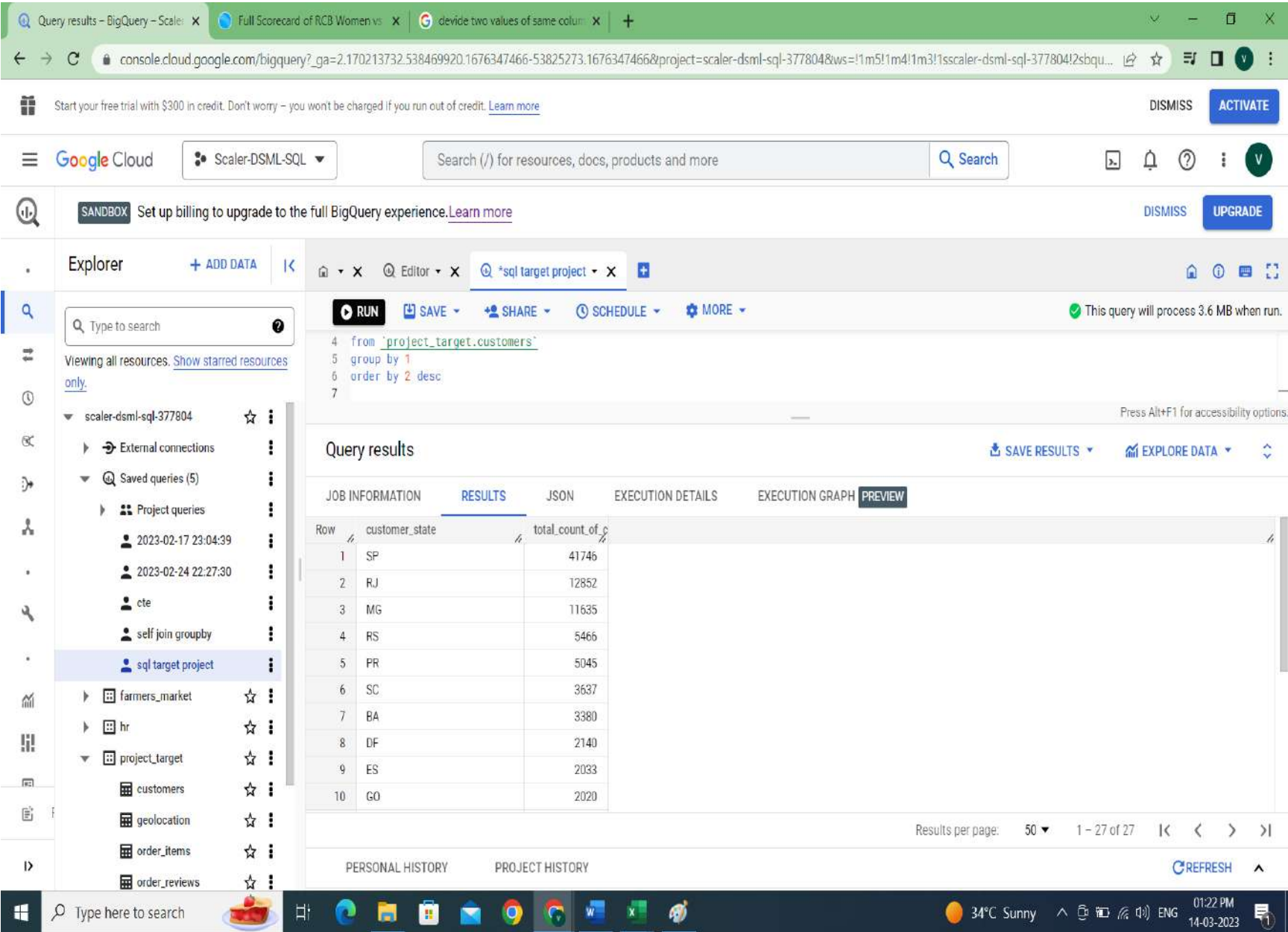
MOM Order Analysis:

- 1. MOM Orders of each States are almost similar. However, sales are coming down in the month of oct-dec which is surprising as this is the Christmas and new year month. Need to explore .

2. Distribution of customers across the states in Brazil

Ans:

```
select
  customer_state,
  count(distinct customer_id) as total_count_of_customers
from `project_target.customers`
group by 1
order by 2 desc
```



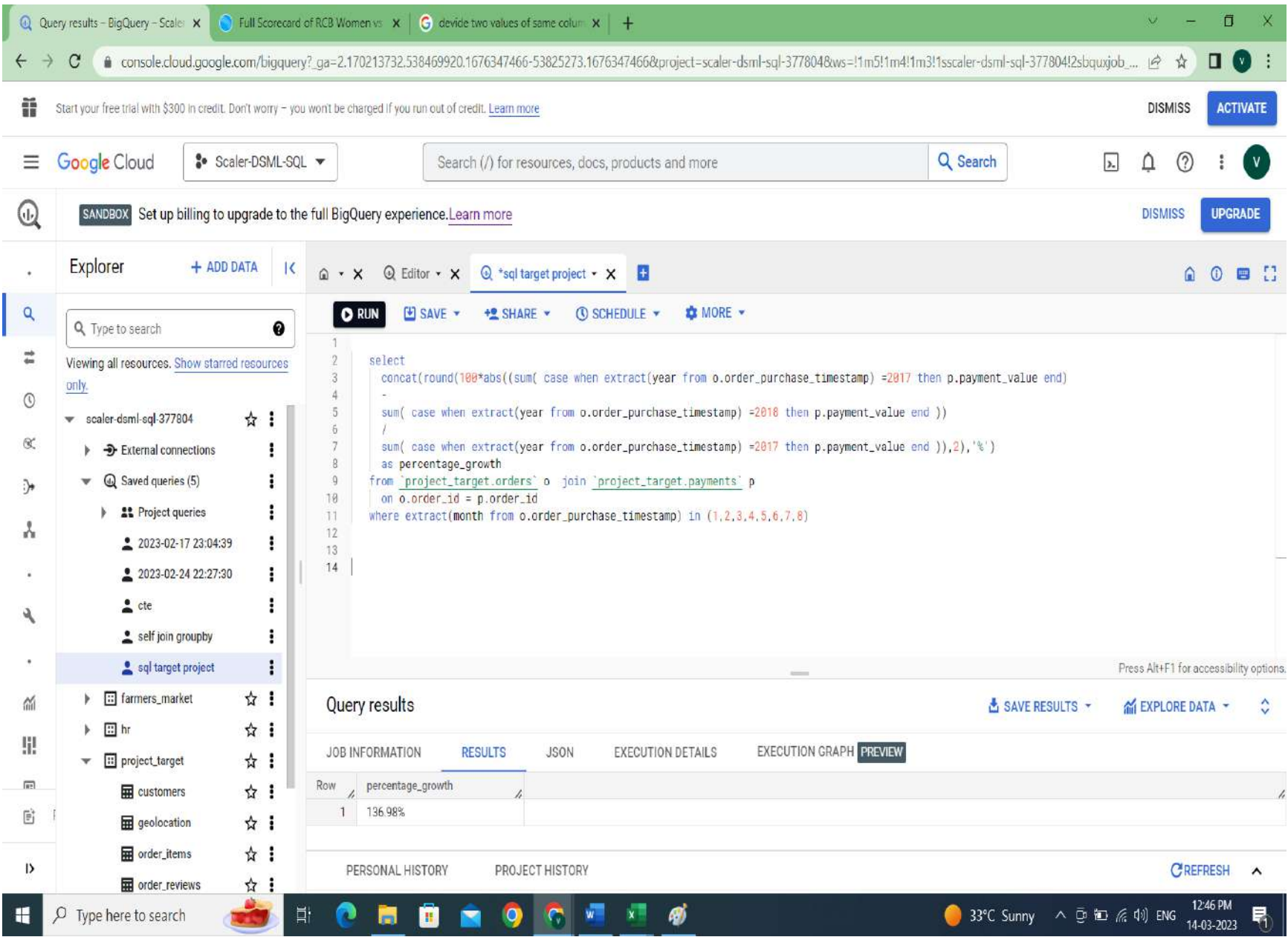
Location Analysis:

- 1. Almost 55% of the customers distributed across two states SP and RJ.

3. 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

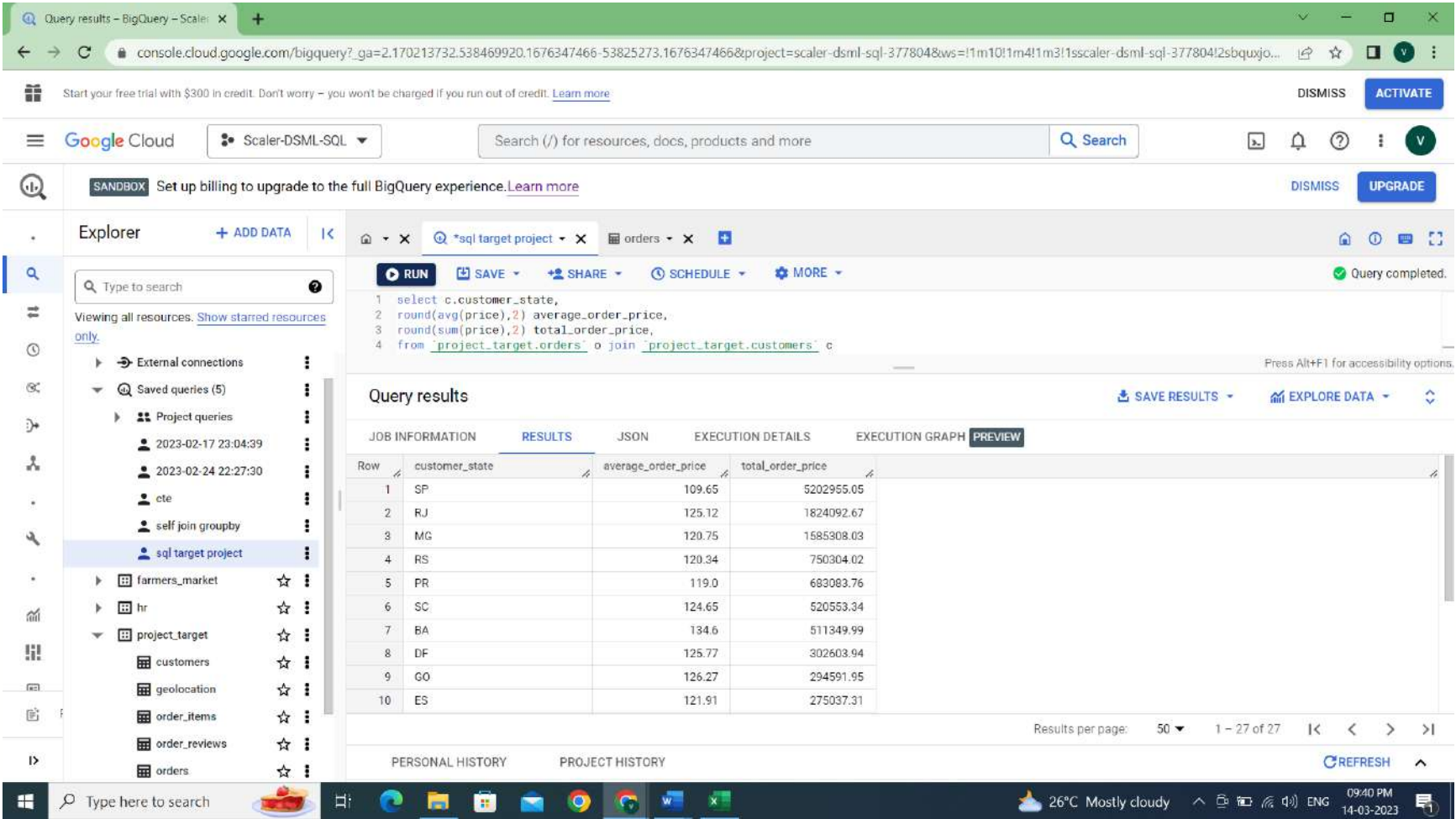
```
select
concat(round(100*abs((sum( case when extract(year from o.order_purchase_timestamp) =2017 then p.payme
nt_value end)
-
sum( case when extract(year from o.order_purchase_timestamp) =2018 then p.payment_value end ))
/
sum( case when extract(year from o.order_purchase_timestamp) =2017 then p.payment_value end )),2),'%')
as percentage_growth
from `project_target.orders` o join `project_target.payments` p
on o.order_id = p.order_id
where extract(month from o.order_purchase_timestamp) in (1,2,3,4,5,6,7,8)
```



2.Mean & Sum of price by customer state

Ans:

```
select c.customer_state,
       round(avg(price),2) average_order_price,
       round(sum(price),2) total_order_price,
from `project_target.orders` o join `project_target.customers` c
  on o.customer_id = c.customer_id
  join `project_target.order_items` ot
  on o.order_id = ot.order_id
group by 1
order by 3 desc
```

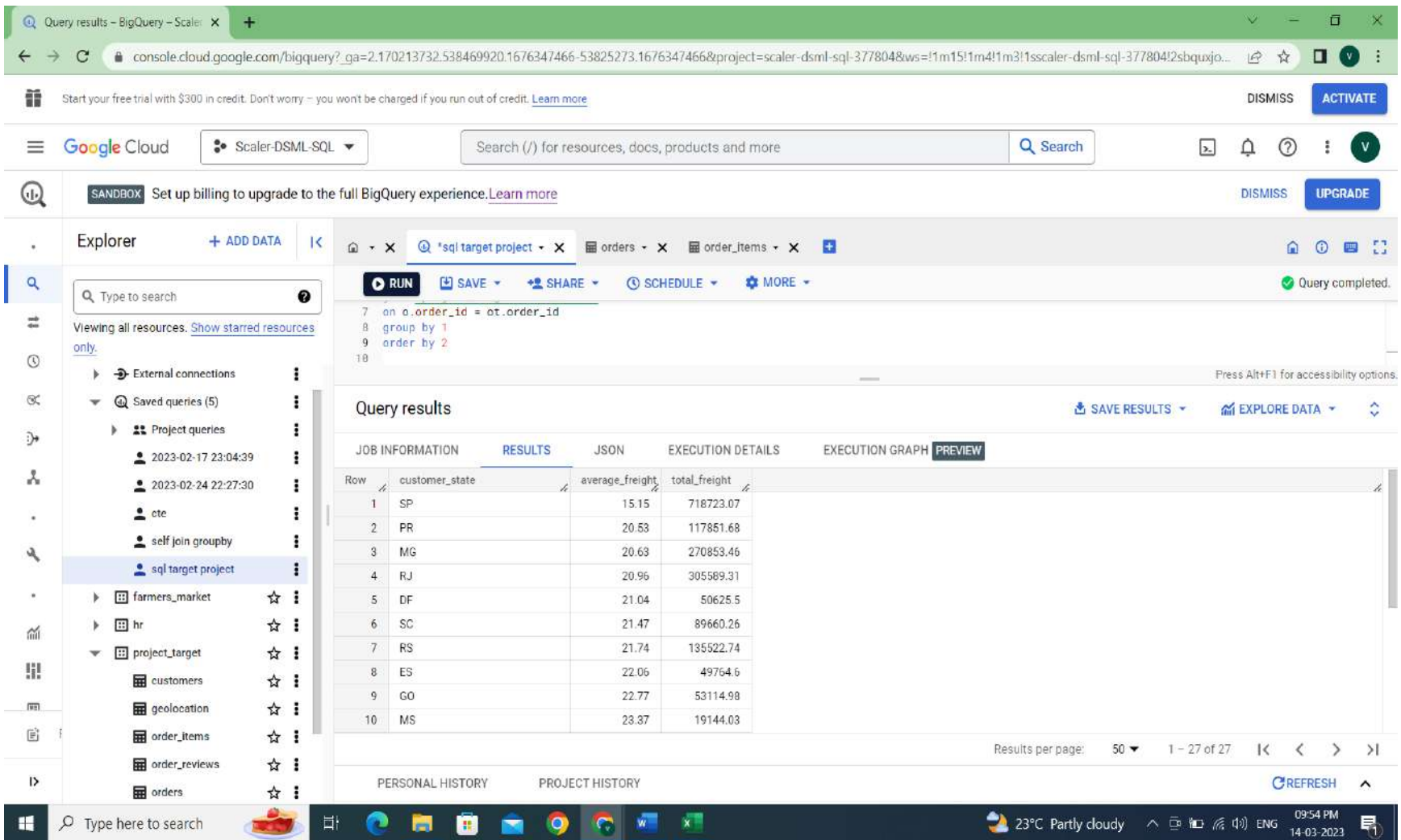


Order Prices analysis:

- 2. We can see that more than 50% of total order price is received from two states SP and RJ. But if we see the average order price for these regions are quite opposite. It shows that high order regions customers’ orders more frequently

2.Mean & Sum of Freight by customer state

```
select c.customer_state,
       round(avg(freight_value),2) average_freight,
       round(sum(freight_value),2) total_freight,
from `project_target.orders` o join `project_target.customers` c
  on o.customer_id = c.customer_id
  join `project_target.order_items` ot
    on o.order_id = ot.order_id
group by 1
order by 2
```



Freight Prices analysis:

- 1. We can see that the average freight price is also lower for states with high order prices and fast delivery.

5. Analysis on sales, freight and delivery time

- 1. Calculate days between purchasing, delivering and estimated delivery
- 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
- 3. Group data by state, take mean of freight_value, time_to_delivery, diff_estimated_delivery

Ans:

```
select
c.customer_state,
round(avg(ot.freight_value),2) as average_freight_value,
round(avg(date_diff(order_delivered_customer_date,order_purchase_timestamp,day)),2) as average_delivery_days,
round(avg(date_diff(order_estimated_delivery_date,order_delivered_customer_date,day)),2) as average_diff_estimated_delivery,
from `project_target.orders` o join `project_target.order_items` ot
on o.order_id = ot.order_id join `project_target.customers` c
on o.customer_id = c.customer_id
where o.order_status = 'delivered'
group by 1
order by 1
```

Query results – BigQuery – Scaler

console.cloud.google.com/bigquery?_ga=2.170213732.538469920.1676347466-53825273.1676347466&project=scaler-dsml-sql-377804&ws=11m10!1m4!1m3!1sscaler-dsml-sql-377804!2sbquxjo...

Start your free trial with \$300 in credit. Don't worry – you won't be charged if you run out of credit. [Learn more](#)

DISMISS

ACTIVATE

Google Cloud

Scaler-DSML-SQL

Search (/) for resources, docs, products and more

Search

🔍

🔔

?

⋮

V

SANDBOX

Set up billing to upgrade to the full BigQuery experience. [Learn more](#)

DISMISS

UPGRADE

Explorer

+ ADD DATA

🔍

Type to search

Viewing all resources. [Show starred resources only.](#)

External connections

Saved queries (5)

Project queries

2023-02-17 23:04:39

2023-02-24 22:27:30

cte

self join groupby

sql target project

farmers_market

hr

project_target

customers

geolocation

order_items

order_reviews

orders

*sql target project

orders

+ RUN

SAVE

SHARE

SCHEDULE

MORE

1 select

2 c.customer_state,

3 round(avg(ot.freight_value),2) as average_freight_value,

4 round(avg(date_diff(order_delivered_customer_date,order_purchase_timestamp,day)),2) as average_delivery_days,

Query results

SAVE RESULTS

EXPLORE DATA

🔍

JOB INFORMATION

RESULTS

JSON

EXECUTION DETAILS

EXECUTION GRAPH

PREVIEW

Row	customer_state	average_freight	average_delivery	average_diff_est
1	AC	40.05	20.33	20.01
2	AL	35.87	23.99	7.98
3	AM	33.31	25.96	18.98
4	AP	34.16	27.75	17.44
5	BA	26.49	18.77	10.12
6	CE	32.73	20.54	10.26
7	DF	21.07	12.5	11.27
8	ES	22.03	15.19	9.77
9	GO	22.56	14.95	11.37
10	MA	38.49	21.2	9.11

Results per page: 50 1 – 27 of 27

PERSONAL HISTORY

PROJECT HISTORY

REFRESH

Type here to search

🔍

📅

📁

📧

🌐

🔌

🔊

🌡️ 26°C Mostly clear

📶

🔌

🌐 ENG

09:08 PM

14-03-2023

🔔

4.Sort the data to get the following:

5.Top 5 states with highest/lowest average freight value - sort in desc/asc limit 5

Ans:

Top 5 with highest average freight –

```
select
c.customer_state,
round(avg(ot.freight_value),2) as average_freight_value,
round(avg(date_diff(order_delivered_customer_date,order_purchase_timestamp,day)),2) as average_delivery_d
ays,
round(avg(date_diff(order_estimated_delivery_date,order_delivered_customer_date,day)),2) as average_diff_es
timated_delivery,
from `project_target.orders` o join `project_target.order_items` ot
on o.order_id = ot.order_id join `project_target.customers` c
on o.customer_id = c.customer_id
group by 1
order by 2 desc
limit 5
```

Query results - BigQuery - Scaler

console.cloud.google.com/bigquery?_ga=2.170213732.538469920.1676347466-53825273.1676347466&project=scaler-dsml-sql-377804&ws=!1m10!1m4!1m3!1sscaler-dsml-sql-377804!2sbquxjo...

Start your free trial with \$300 in credit. Don't worry - you won't be charged if you run out of credit. [Learn more](#)

DISMISS ACTIVATE

Google Cloud

Scaler-DSML-SQL

Search (/) for resources, docs, products and more

Search

DISMISS UPGRADE

Explorer

+ ADD DATA

Q Type to search

Viewing all resources. [Show starred resources only.](#)

External connections

Saved queries (5)

Project queries

2023-02-17 23:04:39

2023-02-24 22:27:30

cte

self join groupby

sql target project

farmers_market

hr

project_target

customers

geolocation

order_items

order_reviews

orders

RUN

SAVE

SHARE

SCHEDULE

MORE

```
1 select
2 c.customer_state,
3 round(avg(ot.freight_value),2) as average_freight_value,
4 round(avg(date_diff(order_delivered_customer_date,order_purchase_timestamp,day)),2) as average_delivery_days,
5 round(avg(date_diff(order_estimated_delivery_date,order_delivered_customer_date,day)),2) as average_diff_estimated_delivery,
6 from `project_target.orders` o join `project_target.order_items` ot
7   on o.order_id = ot.order_id join `project_target.customers` c
8   on o.customer_id = c.customer_id
9 group by 1
10 order by 2 desc
11 limit 5
12
```

Query results

SAVE RESULTS EXPLORE DATA

JOB INFORMATION

RESULTS

JSON

EXECUTION DETAILS

EXECUTION GRAPH

PREVIEW

Row	customer_state	average_freight	average_delivery	average_diff_est
1	RR	42.98	27.83	17.43
2	PB	42.72	20.12	12.15
3	RO	41.07	19.28	19.08
4	AC	40.07	20.33	20.01
5	PI	39.15	18.93	10.68

PERSONAL HISTORY

PROJECT HISTORY

REFRESH

Type here to search

33°C Mostly sunny

05:13 PM 14-03-2023

Top 5 with lowest average freight –

Ans:

```
select
c.customer_state,
round(avg(ot.freight_value),2) as average_freight_value,
round(avg(date_diff(order_delivered_customer_date,order_purchase_timestamp,day)),2) as average_delivery_d
ays,
round(avg(date_diff(order_estimated_delivery_date,order_delivered_customer_date,day)),2) as average_diff_es
timated_delivery,
from `project_target.orders` o join `project_target.order_items` ot
    on o.order_id = ot.order_id join `project_target.customers` c
    on o.customer_id = c.customer_id
group by 1
order by 2
limit 5
```

Query results – BigQuery – Scaler

console.cloud.google.com/bigquery?_ga=2.170213732.538469920.1676347466-53825273.1676347466&project=scaler-dsml-sql-377804&ws=!1m10!1m4!1m3!1sscaler-dsml-sql-377804!2sbquxjo...

Start your free trial with \$300 in credit. Don't worry – you won't be charged if you run out of credit. [Learn more](#)

DISMISS

ACTIVATE

Google Cloud

Scaler-DSML-SQL

Search (/) for resources, docs, products and more

Search

DISMISS

UPGRADE

Explorer

+ ADD DATA

SQL Editor

*sql target project

order_items

RUN

SAVE

SHARE

SCHEDULE

MORE

Query completed.

Type to search

Viewing all resources. [Show starred resources only.](#)

External connections

Saved queries (5)

Project queries

2023-02-17 23:04:39

2023-02-24 22:27:30

cte

self join groupby

sql target project

farmers_market

hr

project_target

customers

geolocation

order_items

order_reviews

orders

1 select

2 c.customer_state,

3 round(avg(ot.freight_value),2) as average_freight_value,

4 round(avg(date_diff(order_delivered_customer_date,order_purchase_timestamp,day)),2) as average_delivery_days,

5 round(avg(date_diff(order_estimated_delivery_date,order_delivered_customer_date,day)),2) as average_diff_estimated_delivery,

6 from `project_target.orders` o join `project_target.order_items` ot

7 on o.order_id = ot.order_id join `project_target.customers` c

8 on o.customer_id = c.customer_id

9 group by 1

10 order by 2

11 limit 5

12

Query results

SAVE RESULTS

EXPLORE DATA

JOB INFORMATION

RESULTS

JSON

EXECUTION DETAILS

EXECUTION GRAPH

PREVIEW

Row	customer_state	average_freight	average_delivery	average_diff_est
1	SP	15.15	8.26	10.27
2	PR	20.53	11.48	12.53
3	MG	20.63	11.52	12.4
4	RJ	20.96	14.69	11.14
5	DF	21.04	12.5	11.27

PERSONAL HISTORY

PROJECT HISTORY

REFRESH

Type here to search

33°C Mostly sunny

05:13 PM

14-03-2023

6.Top 5 states with highest average time to delivery-

Ans:

```
select
c.customer_state,
round(avg(ot.freight_value),2) as average_freight_value,
round(avg(date_diff(order_delivered_customer_date,order_purchase_timestamp,day)),2) as average_delivery_d
ays,
round(avg(date_diff(order_estimated_delivery_date,order_delivered_customer_date,day)),2) as average_diff_es
timated_delivery,
from `project_target.orders` o join `project_target.order_items` ot
on o.order_id = ot.order_id join `project_target.customers` c
on o.customer_id = c.customer_id
group by 1
order by 3 desc
limit 5
```

Query results - BigQuery - Scaler

console.cloud.google.com/bigquery?_ga=2.170213732.538469920.1676347466-53825273.1676347466&project=scaler-dsml-sql-377804&ws=!1m10!1m4!1m3!1ssc...DISMISSACTIVATE

Google CloudScaler-DSML-SQLSearch (/) for resources, docs, products and moreSearchDISMISSUPGRADE

Explorer+ ADD DATA

Type to search

Viewing all resources. Show starred resources only.

External connections

Saved queries (5)

Project queries

2023-02-17 23:04:39

2023-02-24 22:27:30

cte

self join groupby

sql target project

farmers_market

hr

project_target

customers

geolocation

order_items

order_reviews

orders

*sql target project

order_items

RUNSAVESHARESCHEDULEMORE

```
1 select
2 c.customer_state,
3 round(avg(ot.freight_value),2) as average_freight_value,
4 round(avg(date_diff(order_delivered_customer_date,order_purchase_timestamp,day)),2) as average_delivery_days,
5 round(avg(date_diff(order_estimated_delivery_date,order_delivered_customer_date,day)),2) as average_diff_estimated_delivery,
6 from `project_target.orders` o join `project_target.order_items` ot
7 on o.order_id = ot.order_id join `project_target.customers` c
8 on o.customer_id = c.customer_id
9 group by 1
10 order by 3 desc
11 limit 5
12
```

Query resultsSAVE RESULTSEXPLORE DATA

JOB INFORMATIONRESULTSJSONEXECUTION DETAILSEXECUTION GRAPHPREVIEW

Row	customer_state	average_freight	average_delivery	average_diff_est
1	RR	42.98	27.83	17.43
2	AP	34.01	27.75	17.44
3	AM	33.21	25.96	18.98
4	AL	35.84	23.99	7.98
5	PA	35.83	23.3	13.37

PERSONAL HISTORYPROJECT HISTORYREFRESH

Top 5 states with lowest average time to delivery-

Ans:

```
select
c.customer_state,
round(avg(ot.freight_value),2) as average_freight_value,
round(avg(date_diff(order_delivered_customer_date,order_purchase_timestamp,day)),2) as average_delivery_d
ays,
round(avg(date_diff(order_estimated_delivery_date,order_delivered_customer_date,day)),2) as average_diff_es
timated_delivery,
from `project_target.orders` o join `project_target.order_items` ot
on o.order_id = ot.order_id join `project_target.customers` c
on o.customer_id = c.customer_id
group by 1
order by 3
limit 5
```

Query results - BigQuery - Scaler

console.cloud.google.com/bigquery?_ga=2.170213732.538469920.1676347466-53825273.1676347466&project=scaler-dsml-sql-377804&ws=1m10!1m4!1m3!1sscaler-dsml-sql-377804!2sbqxjo...

Start your free trial with \$300 in credit. Don't worry - you won't be charged if you run out of credit. [Learn more](#)

DISMISS

ACTIVATE

Google Cloud

Scaler-DSML-SQL

Search (/) for resources, docs, products and more

Search

DISMISS

UPGRADE

SANDBOX

Set up billing to upgrade to the full BigQuery experience. [Learn more](#)

DISMISS

UPGRADE

Explorer

+ ADD DATA

Type to search

Viewing all resources. [Show starred resources only.](#)

External connections

Saved queries (5)

Project queries

2023-02-17 23:04:39

2023-02-24 22:27:30

cte

self join groupby

sql target project

farmers_market

hr

project_target

customers

geolocation

order_items

order_reviews

orders

RUN

SAVE

SHARE

SCHEDULE

MORE

This query will process 16.82 MB when run.

1 select

2 c.customer_state,

3 round(avg(ot.freight_value),2) as average_freight_value,

4 round(avg(date_diff(order_delivered_customer_date,order_purchase_timestamp,day)),2) as average_delivery_days,

5 round(avg(date_diff(order_estimated_delivery_date,order_delivered_customer_date,day)),2) as average_diff_estimated_delivery,

6 from `project_target.orders` o join `project_target.order_items` ot

7 on o.order_id = ot.order_id join `project_target.customers` c

8 on o.customer_id = c.customer_id

9 group by 1

10 order by 3

11 limit 5

12

Query results

SAVE RESULTS

EXPLORE DATA

JOB INFORMATION

RESULTS

JSON

EXECUTION DETAILS

EXECUTION GRAPH

PREVIEW

Row	customer_state	average_freight	average_delivery	average_diff_est
1	SP	15.15	8.26	10.27
2	PR	20.53	11.48	12.53
3	MG	20.63	11.52	12.4
4	DF	21.04	12.5	11.27
5	SC	21.47	14.52	10.67

PERSONAL HISTORY

PROJECT HISTORY

REFRESH

Type here to search

33°C Mostly sunny

05:16 PM

14-03-2023

Analysis on Delivery time:

- 1. We see a strong relation between the number of sellers in the region and the average order delivery speed.
In the states like SP, PR, MG the average delivery time is less than 12 days and we can observe also that the seller count in these states is also high.
We can see the states with delivery time more than 25 days like RR, AP, AM has either only 1 seller or no seller in these states.

7. Top 5 states where delivery is really fast compared to estimated date

```
select
c.customer_state,
round(avg(date_diff(order_estimated_delivery_date,order_delivered_customer_date,day)),2)
as avg_delivery_estimate_comparision
from `project_target.orders` o
join `project_target.customers` c
on o.customer_id = c.customer_id
group by 1
order by 2
limit 5
```

Query results - BigQuery - Scaler

console.cloud.google.com/bigquery?_ga=2.170213732.538469920.1676347466-53825273.1676347466&project=scaler-dsml-sql-377804&ws=!1m10!1m4!1m3!1sscaler-dsml-sql-377804!2sbquxjo...

Start your free trial with \$300 in credit. Don't worry - you won't be charged if you run out of credit. [Learn more](#)

DISMISS

ACTIVATE

Google Cloud

Scaler-DSML-SQL

Search (/) for resources, docs, products and more

Search

DISMISS

UPGRADE

SANDBOX

Set up billing to upgrade to the full BigQuery experience. [Learn more](#)

DISMISS

UPGRADE

Explorer

+ ADD DATA

Viewing all resources. [Show starred resources](#) only.

External connections

Saved queries (5)

Project queries

2023-02-17 23:04:39

2023-02-24 22:27:30

cte

self join groupby

sql target project

farmers_market

hr

project_target

customers

geolocation

order_items

order_reviews

orders

*sql target project

order_items

RUN

SAVE

SHARE

SCHEDULE

MORE

This query will process 8.32 MB when run.

1 select

2 c.customer_state,

3 round(avg(date_diff(order_estimated_delivery_date,order_delivered_customer_date,day)),2) as avg_delivery_estimate_comparision

4 from `project_target.orders` o

5 join `project_target.customers` c

6 on o.customer_id = c.customer_id

7 group by 1

8 order by 2

9 limit 5

10

Query results

SAVE RESULTS

EXPLORE DATA

JOB INFORMATION

RESULTS

JSON

EXECUTION DETAILS

EXECUTION GRAPH

PREVIEW

Row	customer_state	avg_delivery_est
1	AL	7.95
2	MA	8.77
3	SE	9.17
4	ES	9.62
5	BA	9.93

PERSONAL HISTORY

PROJECT HISTORY

REFRESH

Type here to search

33°C Mostly sunny

05:08 PM

14-03-2023

Top 5 states where delivery is not so fast compared to estimated date

```
select
c.customer_state,
round(avg(date_diff(order_estimated_delivery_date,order_delivered_customer_date,day)),2) as avg_delivery_es
timate_comparision
from `project_target.orders` o
join `project_target.customers` c
on o.customer_id = c.customer_id
group by 1
order by 2 desc
limit 5
```

Query results - BigQuery - Scaler

console.cloud.google.com/bigquery?_ga=2.170213732.538469920.1676347466-53825273.1676347466&project=scaler-dsml-sql-377804&ws=!1m10!1m4!1m3!1sscaler-dsml-sql-377804!2sbqxjo...

Start your free trial with \$300 in credit. Don't worry - you won't be charged if you run out of credit. [Learn more](#)

DISMISS

ACTIVATE

Google Cloud

Scaler-DSML-SQL

Search (/) for resources, docs, products and more

Search

DISMISS

UPGRADE

SANDBOX

Set up billing to upgrade to the full BigQuery experience. [Learn more](#)

DISMISS

UPGRADE

Explorer

+ ADD DATA

Q Type to search

Viewing all resources. [Show starred resources only.](#)

External connections

Saved queries (5)

Project queries

2023-02-17 23:04:39

2023-02-24 22:27:30

cte

self join groupby

sql target project

farmers_market

hr

project_target

customers

geolocation

order_items

order_reviews

orders

Q *sql target project

order_items

RUN

SAVE

SHARE

SCHEDULE

MORE

1 select

2 c.customer_state,

3 round(avg(date_diff(order_estimated_delivery_date,order_delivered_customer_date,day)),2) as avg_delivery_estimate_comparision

4 from `project_target.orders` o

5 join `project_target.customers` c

6 on o.customer_id = c.customer_id

7 group by 1

8 order by 2 desc

9 limit 5

10

Query results

SAVE RESULTS

EXPLORE DATA

JOB INFORMATION

RESULTS

JSON

EXECUTION DETAILS

EXECUTION GRAPH

PREVIEW

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

PERSONAL HISTORY

PROJECT HISTORY

REFRESH

Type here to search

33°C Mostly sunny

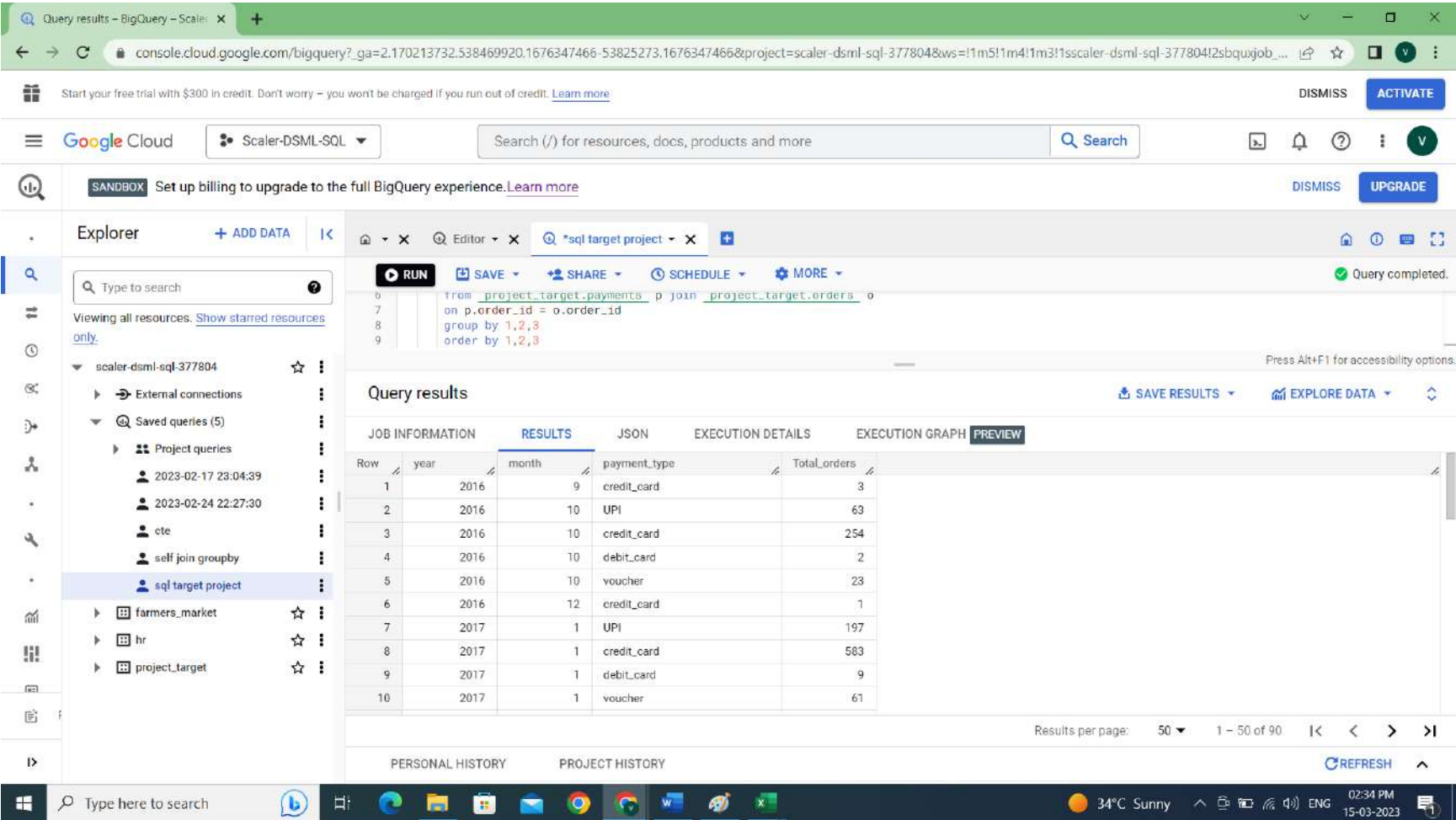
05:07 PM

14-03-2023

6. Payment type analysis:

1. Month over Month count of orders for 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(o.order_id) as Total_orders
from `project_target.payments` p join `project_target.orders` o
on p.order_id = o.order_id
group by 1,2,3
order by 1,2,3
```



Payment type analysis:

- 1. There is slight increase in the use of debit card payment method
- 2. Credit card and UPI is the First and second highest mode of payment respectively and trend is growing very much
- 3. Voucher payment mode is decreasing.

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

select
payment_installments,
count(order_id) as total_orders
from `project_target.payments`
group by 1

Query results - BigQuery - Scaler

console.cloud.google.com/bigquery?_ga=2.170213732.538469920.1676347466-53825273.1676347466&project=scaler-dsml-sql-377804&ws=11m511m411m311ssc...DISMISSACTIVATE

Google CloudScaler-DSML-SQLSearch (/) for resources, docs, products and moreSearchDISMISSUPGRADE

Explorer+ ADD DATA

Type to search

Viewing all resources. Show starred resources only.

scaler-dsml-sql-377804

External connections

Saved queries (5)

Project queries

2023-02-17 23:04:39

2023-02-24 22:27:30

cte

self join groupby

sql target project

farmers_market

hr

project_target

Editor *sql target project

RUNSAVESHARESCHEDULEMORE

This query will process 4.16 MB when run.

3count(order_id) as total_orders

4from `project_target.payments`

5group by 1

6

Query results

SAVE RESULTSEXPLORE DATA

JOB INFORMATIONRESULTSJSONEXECUTION DETAILSEXECUTION GRAPHPREVIEW

Row	payment_installments	total_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

Results per page: 501 - 24 of 24REFRESH

PERSONAL HISTORYPROJECT HISTORY

Type here to search

34°C Sunny02:35 PM15-03-2023

7. Actionable Insights

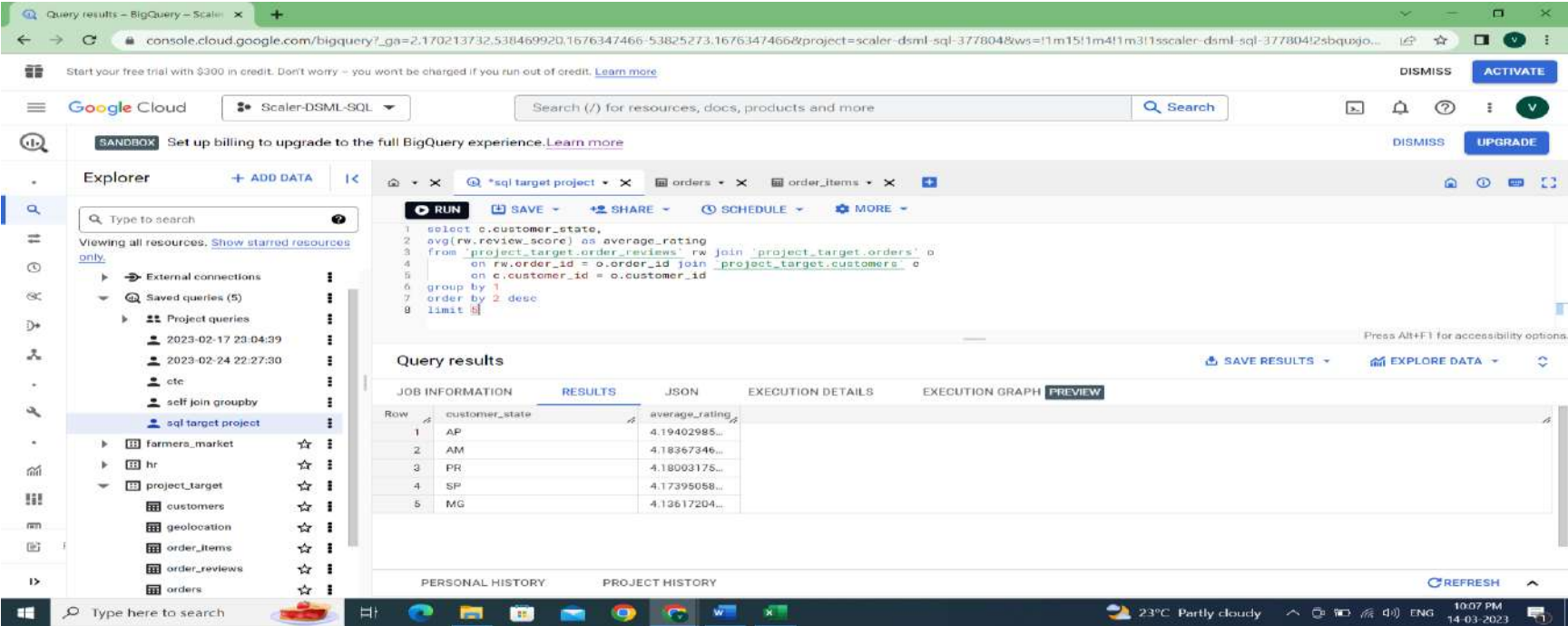
1. Time period Analysis:

- The data set has orders for **2 yrs. (Sep 2016-Oct 2018)** across Brazil from **27 states** which includes **4119 cities**.
-

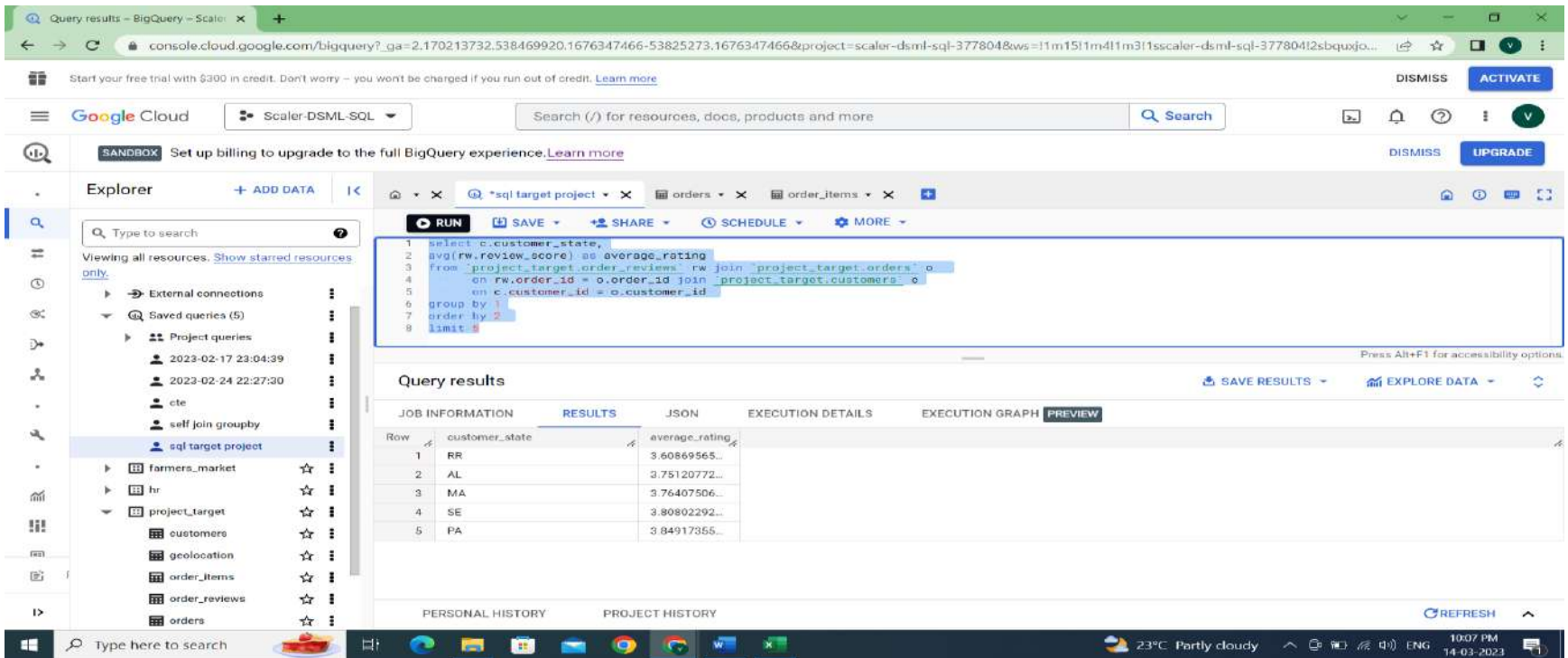
2. Customer rating analysis by state:

- Most active regions in terms of total order count has good review by customer like PR,SP,MG.
 - The delivery time is negatively correlated with the rating. RR,AL,PA states. Longer the delivery time, the lower the score.
 - People are commenting less if they are moderately satisfied.
 - There are more negative reviews (1) than the average review.(2 and 3)
 - Mostly the positive reviews is related with the time of delivery.
- Average rating for delivery day upto 20 days is 4.2*.
Average rating for delivery day 20 to 40 days is 2.93*.
Average rating for delivery day above 40 days is 2.33*.

```
select
  c.customer_state,
  avg(rw.review_score) as average_rating
from `project_target.order_reviews` rw join `project_target.orders` o
  on rw.order_id = o.order_id join `project_target.customers` c
  on c.customer_id = o.customer_id
group by 1
order by 2 desc
limit 5
```



```
select c.customer_state,
avg(rw.review_score) as average_rating
from `project_target.order_reviews` rw join `project_target.orders` o
  on rw.order_id = o.order_id join `project_target.customers` c
  on c.customer_id = o.customer_id
group by 1
order by 2
limit 5
```

3. Trend Analysis:

- We see the year and month wise data with order count, overall, there is a growing trend.
- We can see that the sales count increase from month of **Jan till March**. This could be due to **carnival festival** held at Brazil every year around **last week of Feb till 1st Week of March**.

4. Purchase time Analysis:

- We can see from the data that, majority of purchase happens on afternoon time(13:00 hrs. – 18:00 hrs.)
- Morning (07:00 hrs. – 12:00 hrs.) and night purchases (19:00-23:00) are almost similar.
- Least purchase happens at Dawn (0:00 hrs. – 06:00 hrs.)

5. Order Prices analysis:

- We can see that more than 50% of total order price is received from two states SP and RJ. But if we see the average order price for these regions are quite opposite. It shows that high order regions customers’ orders more frequently.

6. Freight Prices analysis:

- We can see that the average freight price is also lower for states with high order prices and fast delivery.

7. Analysis on Delivery time:

- We see a strong relation between the number of sellers in the region and the average order delivery speed. In the states like SP, PR, MG the average delivery time is less than 12 days and we can observe also that the seller count in these states is also high. We can see the states with delivery time more than 25 days like RR, AP, AM has either only 1 seller or no seller in these states.
- Company seems to be doing well in terms of delivery time since most of the orders are getting delivered way before they are expected.

8. Payment type analysis:

- There is slight increase in the use of debit card payment method
- Credit card and UPI is the First and second highest mode of payment respectively and trend is growing very much
- Voucher payment mode is decreasing.
- People prefer digital mode of payment and credit card is preferred the most.

9. MOM Order Analysis:

- MOM Orders of each States are almost similar. However, sales are coming down in the month of oct-dec which is surprising as this is the Christmas and new year month. Need to explore.

10. Location Analysis:

- Almost 55% of the customers distributed across two states SP and RJ.

11. Top Selling Products:

- Top 3 selling product names are: **Bed table bath(1115), Health beauty(9670) and Sports leisure(8641).**
- Least 3 selling product names are : **Insurance services(2), Fashion children clothing(8) and PC gamer(9).**

8. Recommendations:

- Can exploit more festive seasons like Christmas and new year for more sales.
- Almost 55% of the customers distributed across two states SP and RJ. Customers are not proportionally distributed. Need to explore for getting more customers from other states as well.
- More sellers can be acquired for more sales. We can see that there is a strong relation between number of sellers and order delivery speed which is related to more sales.
- The consequence of late delivery is that the customers are likely to give lower reviews.
- Should start focussing product recommended system for customers. Depending upon the purchase made, geographical location etc.
- Since purchasing is highest in the afternoon time, more promotional ads, mails, messages can be sent for more visibility.