# **SQL Project –E-Commerce Sales Analysis Queries**

**Your client, a Brazilian E-commerce store, wants you to help them understand how their sales trend across different states over the years and why?**

**For this analysis, you are given the data for the time period 2016-2018**

**A. Create the different metrics like Sales, customer acquisitions, total no. of orders for each Year across the different states they serve.**

**Query: 1) for number of orders across each state throughout the years.**

SELECT \* FROM (

SELECT customer\_state AS State, year(order\_purchase\_timestamp) AS Year, COUNT(order\_id) AS No\_of\_orders FROM olist\_orders\_dataset AS a

JOIN olist\_customers\_dataset AS b

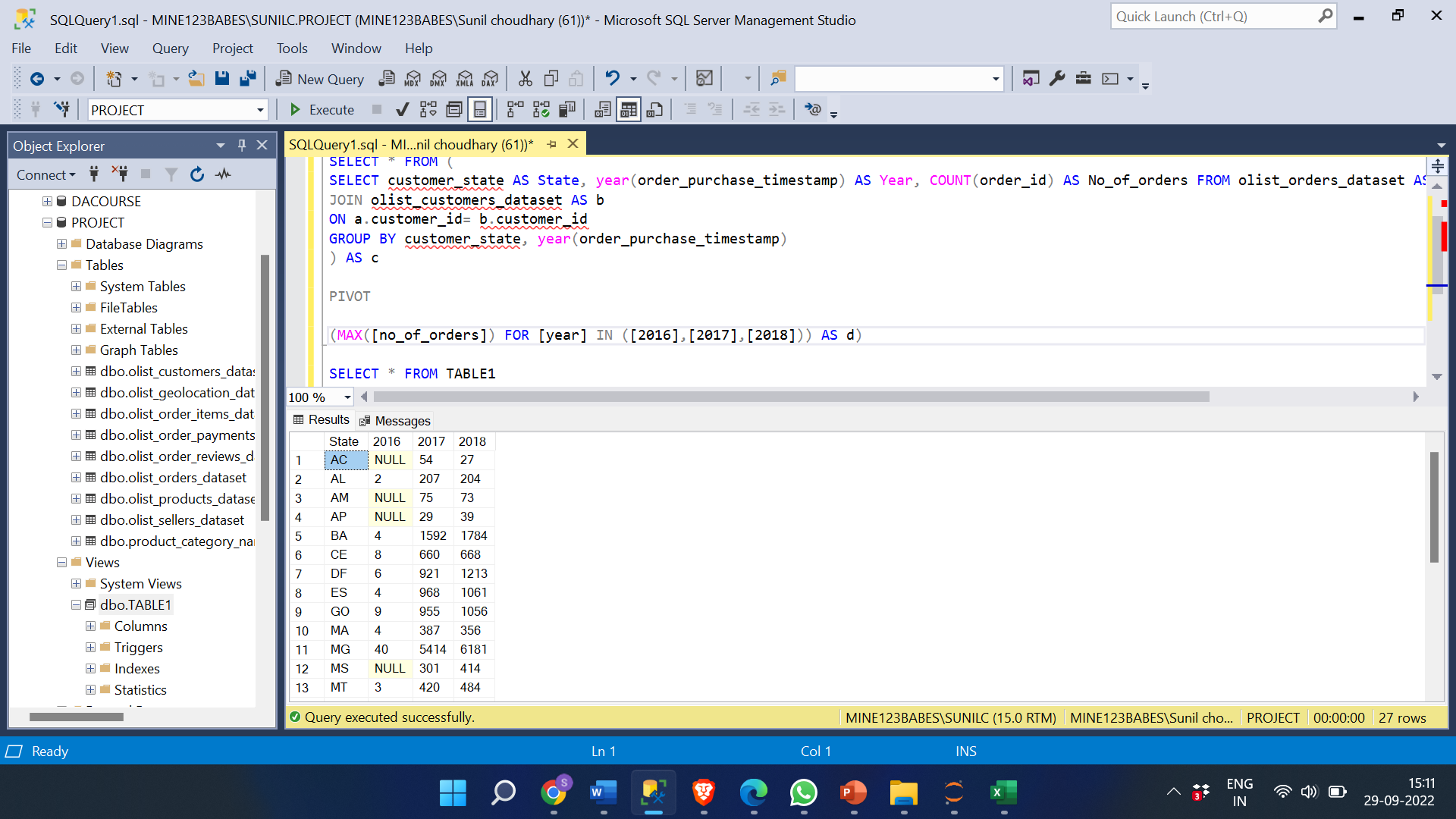
ON a.customer\_id= b.customer\_id

GROUP BY customer\_state, year(order\_purchase\_timestamp)

) AS c

PIVOT

(MAX([no\_of\_orders]) FOR [year] IN ([2016],[2017],[2018])) AS d



**Query: 2) for number of customers across each state throughout the years.**

SELECT \* FROM (

SELECT customer\_state AS State, Year(order\_purchase\_timestamp) AS Year, COUNT(b.customer\_id) AS No\_of\_customers FROM olist\_orders\_dataset AS a

JOIN olist\_customers\_dataset AS b

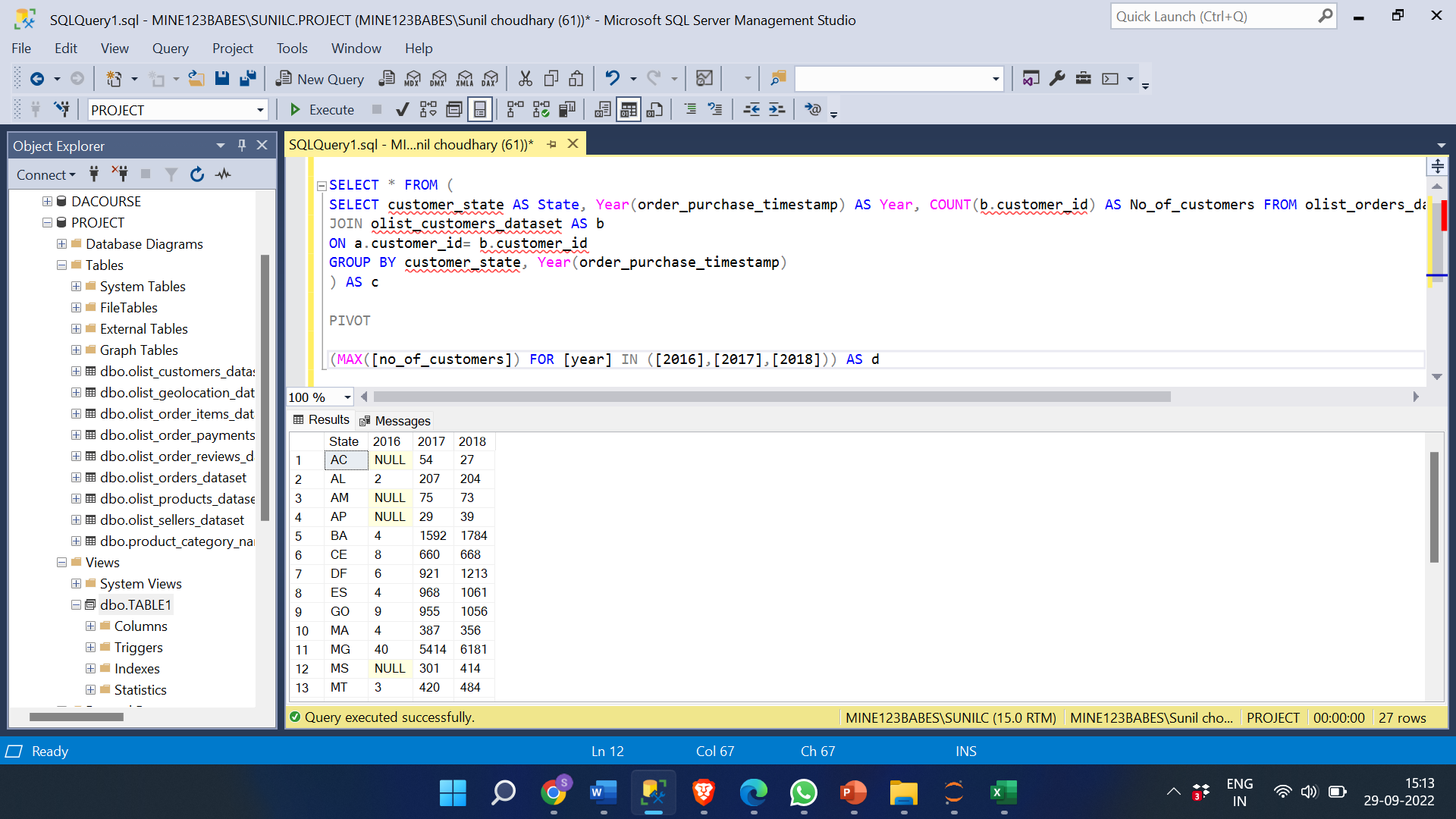
ON a.customer\_id= b.customer\_id

GROUP BY customer\_state, Year(order\_purchase\_timestamp)

) AS c

PIVOT

(MAX([no\_of\_customers]) FOR [year] IN ([2016],[2017],[2018])) AS d



**QUERY: 3) For Sales Across Each State Throughout The Years.**

SELECT \* FROM (

SELECT Customer\_State AS State, Year(Order\_Purchase\_Timestamp) AS Year, SUM(payment\_value) AS Sales

FROM olist\_orders\_dataset AS a

JOIN olist\_order\_payments\_dataset AS b

ON a.order\_id = b.order\_id

JOIN olist\_customers\_dataset AS c

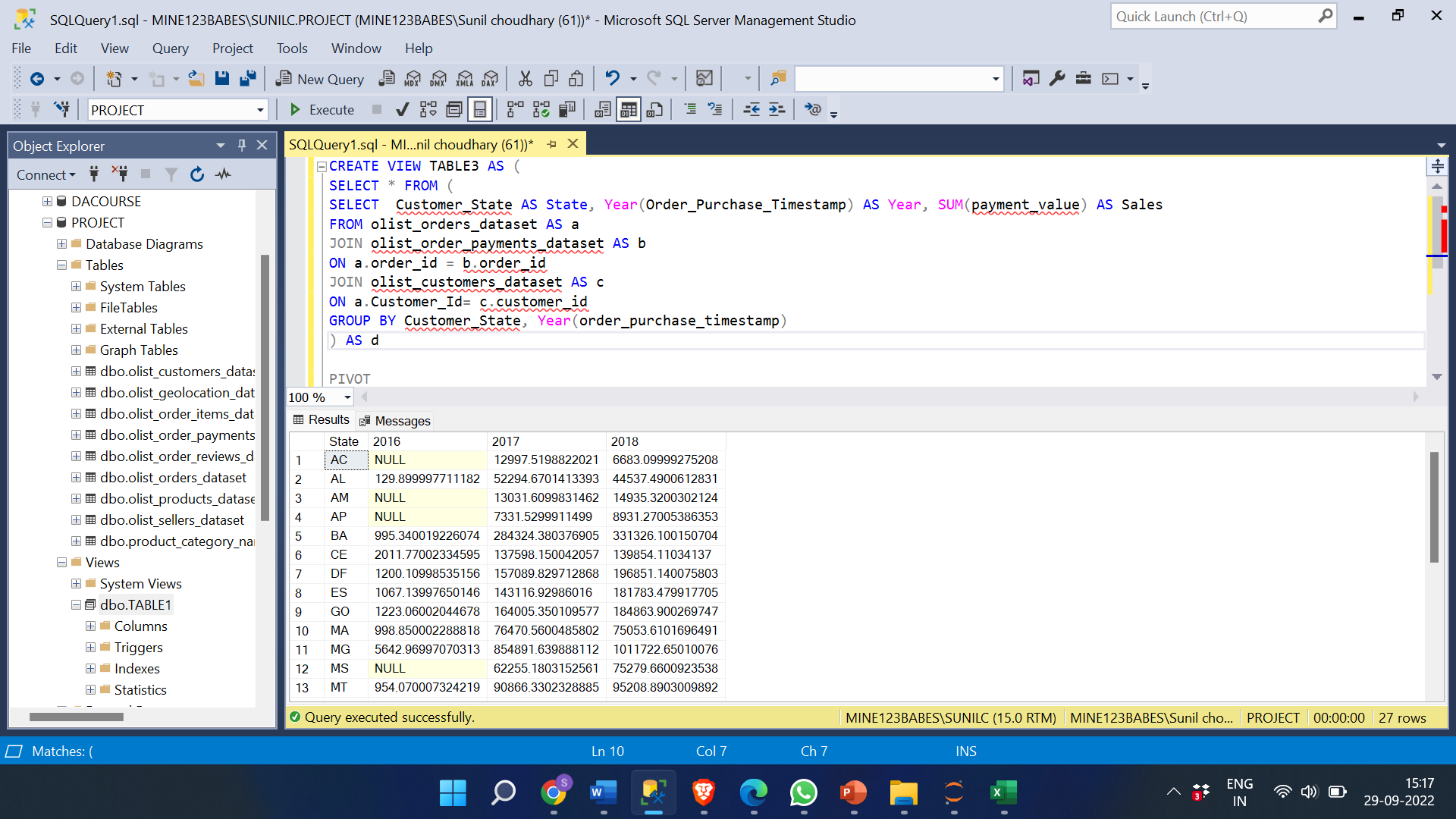
ON a.Customer\_Id= c.customer\_id

GROUP BY Customer\_State, Year(order\_purchase\_timestamp)

) AS d

PIVOT

(MAX([Sales]) FOR [year] IN ([2016],[2017],[2018])) AS f



**INSIGHT: SP has the highest number of customers, number of orders, sales and RR has the lowest sales, orders and customers.**

**Q. Does all the metrics show similar trends or is there any disparity amongst each of them?**

**Ans.**

**1.** Yes, all the metrics show similar trends, the orders and sale from that state is directly proportional to the number of customers.

2. There are few exceptions some states don’t follow trends, most states follows increase in number of orders, sales, and customers.

**b. Using the above metrics, identify the top 2 States which show**

**1. Declining trend over the years**

**Ans.**

**The 3 states which showed Declining trend over the years.**

**They are RO, AC and AM**

* **We used the Number of orders table for this analysis.**
* **I created a view named XYZ containing all the orders from all the states.**

**Query:**

SELECT \* FROM (

SELECT customer\_state AS state, Year(order\_purchase\_timestamp) AS Year, COUNT(order\_id) AS No\_of\_orders FROM olist\_orders\_dataset AS a

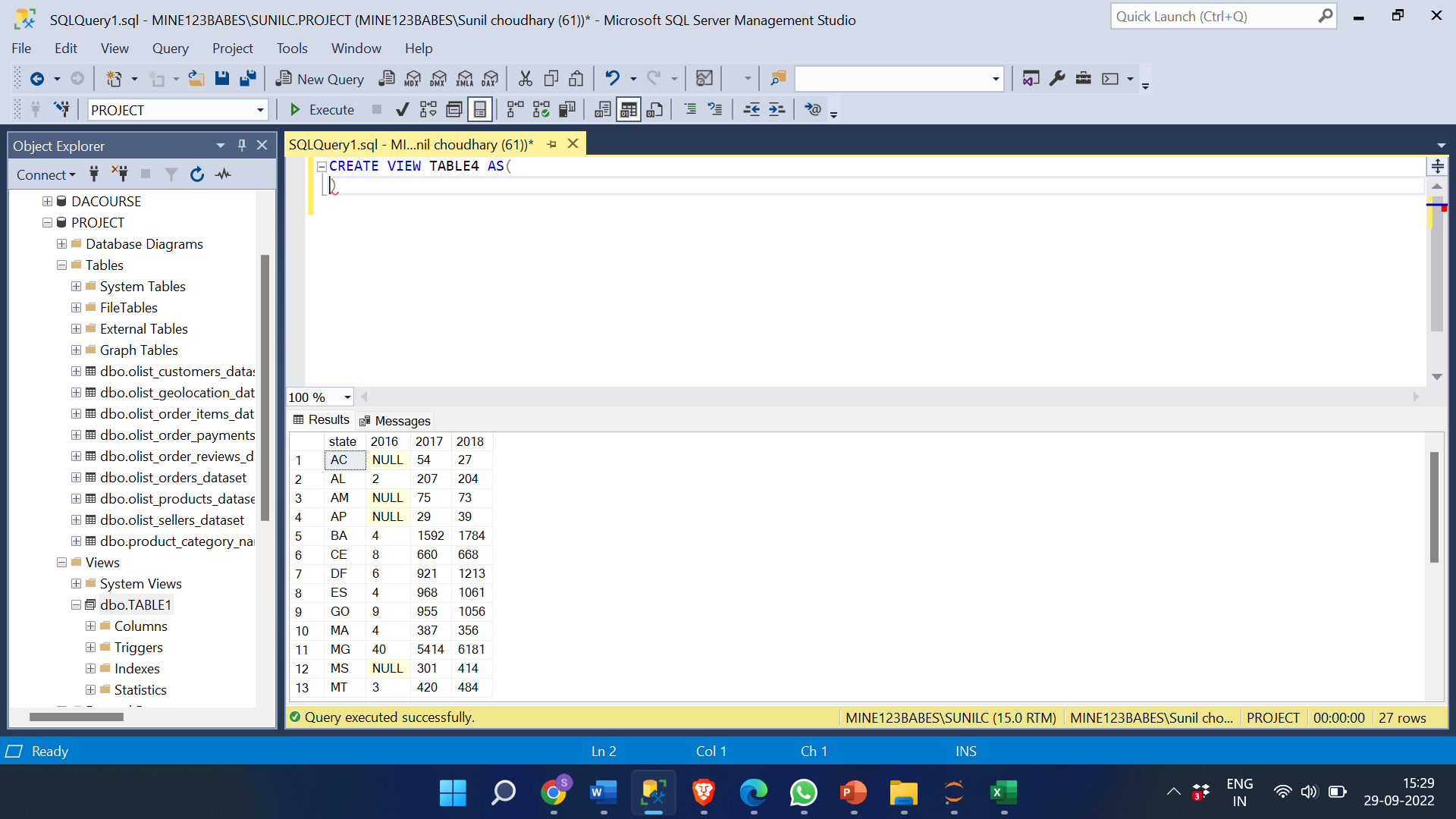
JOIN olist\_customers\_dataset AS b

ON a.customer\_id= b.customer\_id

GROUP BY customer\_state, Year(order\_purchase\_timestamp)) AS d

PIVOT

(MAX([no\_of\_orders]) FOR [year] IN ([2016],[2017],[2018])) AS f



**Next step was to find out the states which show declining trend among all the states.**

**Query:**

SELECT state,[2016], [2017], [2018],dec\_trend\_16\_18,dec\_trend\_17\_18,DENSE\_RANK() OVER (ORDER BY decider DESC ,dec\_trend\_17\_18 ) AS Ranking FROM (

SELECT \*,(CASE

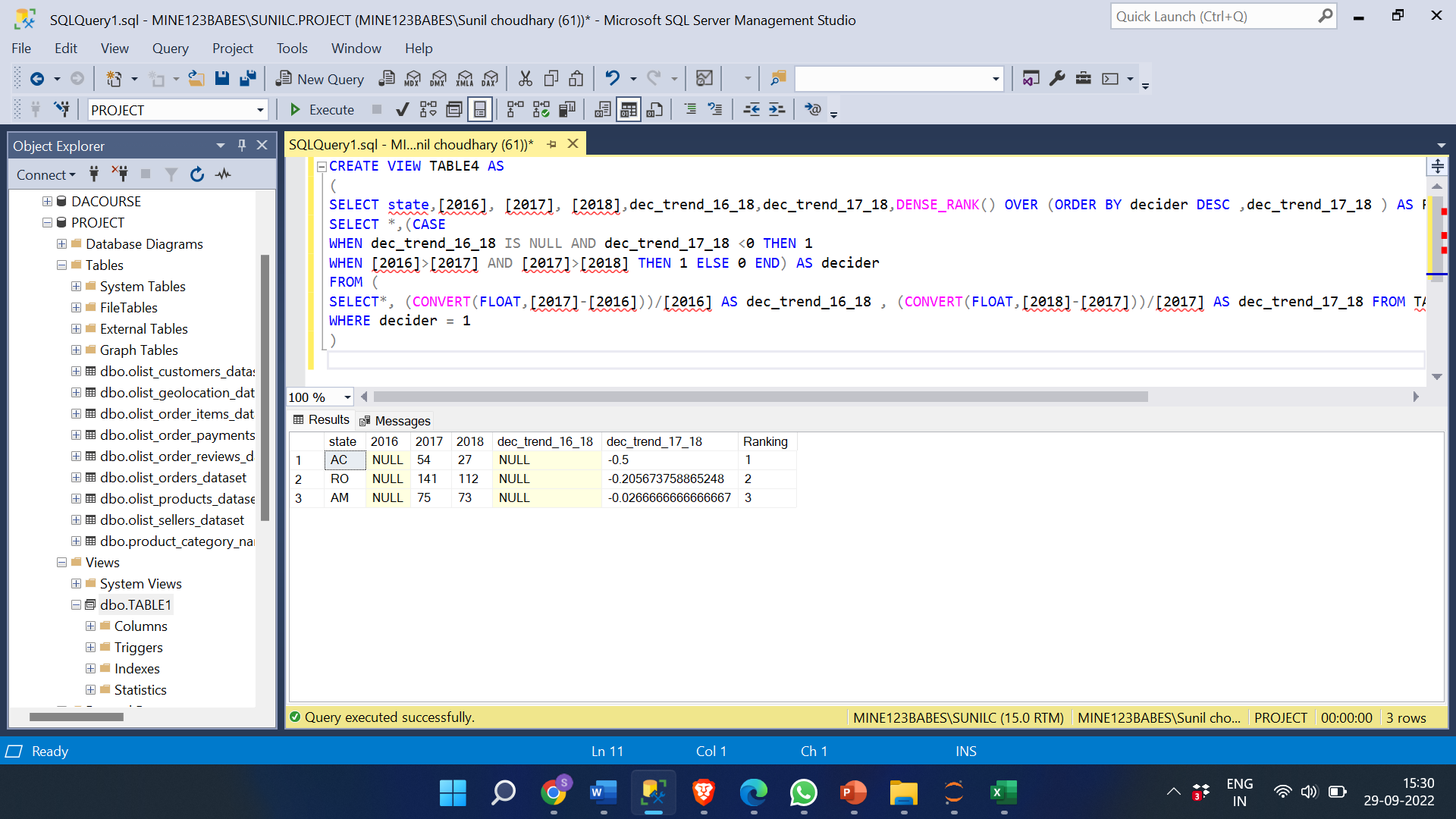
WHEN dec\_trend\_16\_18 IS NULL AND dec\_trend\_17\_18 <0 THEN 1

WHEN [2016]>[2017] AND [2017]>[2018] THEN 1 ELSE 0 END) AS decider

FROM (

SELECT\*, (CONVERT(FLOAT,[2017]-[2016]))/[2016] AS dec\_trend\_16\_18 , (CONVERT(FLOAT,[2018]-[2017]))/[2017] AS dec\_trend\_17\_18 FROM XYZ)AS d) AS f

WHERE decider = 1



**Out of total of 27 states these 3 states showed decline trend and the top 2 states with most decline was calculated through decreasing percentage from 2016 onwards.**

**- The top 2 states were: RO and AC with most decline percentage.**

**Query:**

SELECT TOP 2 State,[2016], [2017], [2018],dec\_trend\_16\_18,dec\_trend\_17\_18,DENSE\_RANK() OVER (ORDER BY decider DESC ,dec\_trend\_17\_18 )AS Ranking FROM (

SELECT \*,

(CASE

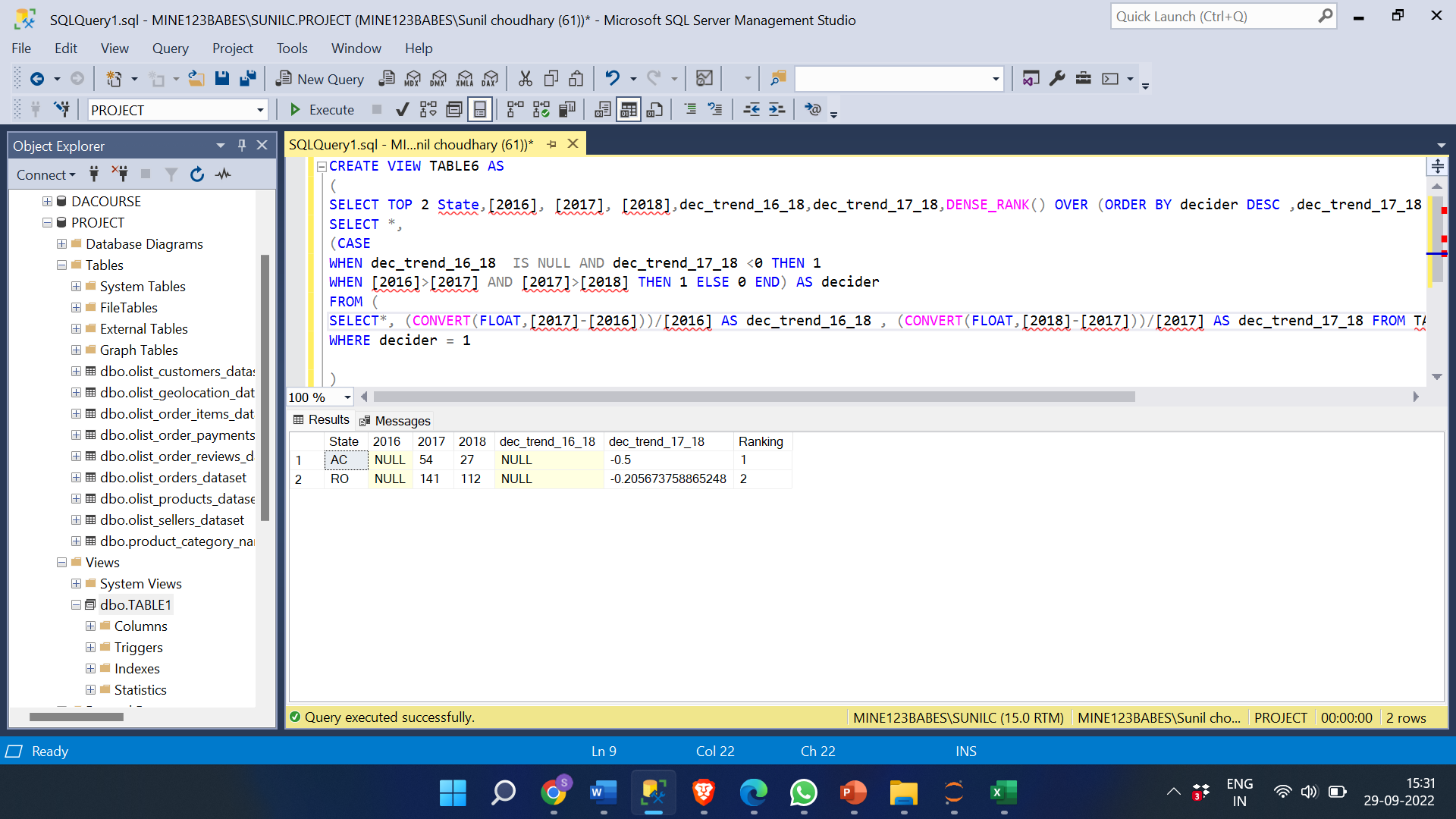
WHEN dec\_trend\_16\_18 IS NULL AND dec\_trend\_17\_18 <0 THEN 1

WHEN [2016]>[2017] AND [2017]>[2018] THEN 1 ELSE 0 END) AS decider

FROM (

SELECT\*, (CONVERT(FLOAT,[2017]-[2016]))/[2016] AS dec\_trend\_16\_18 , (CONVERT(FLOAT,[2018]-[2017]))/[2017] AS dec\_trend\_17\_18 FROM TABLE3)AS dd) AS f

WHERE decider = 1



**Increasing trend over the years.**

**the states with increasing trend over the years were calculated using increasing percentage and then among them the top two states with most increase were calculated.**

SELECT TOP 2 \*, DENSE\_RANK () OVER ( ORDER BY incre\_trend\_16\_17 DESC, incre\_trend\_17\_18 DESC) AS ranking

FROM (

SELECT State, [2016], [2017], [2018], CONVERT (FLOAT,([2017]-[2016]))/[2016] AS incre\_trend\_16\_17, CONVERT (FLOAT,([2018]-[2017]))/[2017] AS incre\_trend\_17\_18

FROM (

SELECT \* FROM (

SELECT \*, (CASE

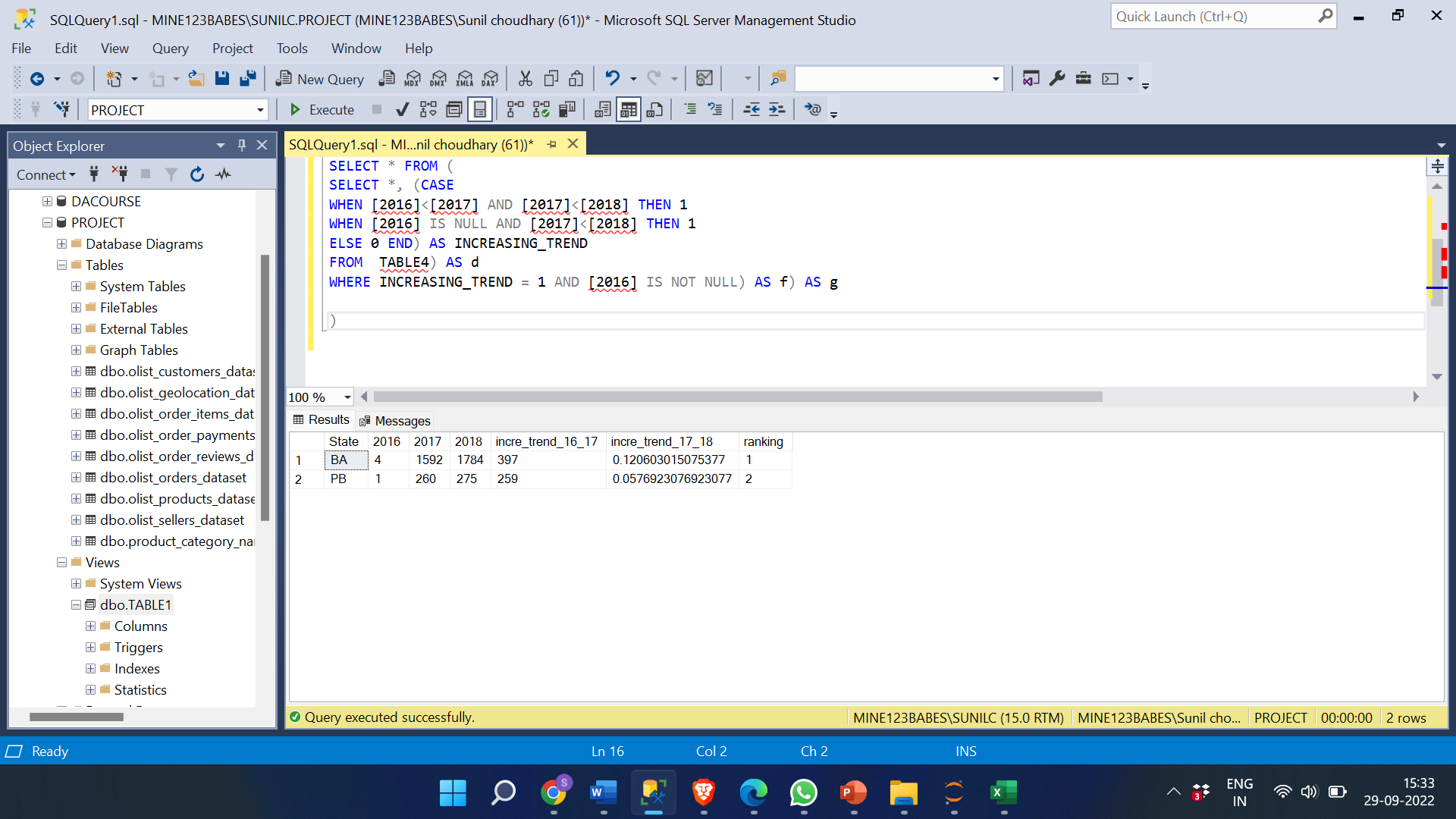
WHEN [2016]<[2017] AND [2017]<[2018] THEN 1

WHEN [2016] IS NULL AND [2017]<[2018] THEN 1

ELSE 0 END) AS INCREASING\_TREND

FROM XYZ ) AS d

WHERE INCREASING\_TREND = 1 AND [2016] IS NOT NULL) AS f) AS g



**INSIGHTS:**

* **Majority of the states show increasing trend in the number of orders and from them BA and PB showed most increase in number of orders over the years.**
* **States with less sales is due to smaller number of people in them which led to less orders and customers.**
* **Geographical location also affects the sales as in urban population prefer online shopping more when compared to rural areas.**
* **Due to less shippers shipping the orders can control the delivery charges and affect the sales.**
* **The trust on online websites used for shopping also affect the sales.**
* **The sales can be affected by the past orders, past experience, marketing, etc.,.**
* **Poor forecasting can affect the sales, the stock in the store leading to customers orders cannot be fullfilled.**
* **Post sell services provided by the companies and the experience affects the sales.**

**The reasons for increase in sales over the years are:**

* **Introducing new products and services over years.**
* **Providing huge benefits and discounts on regular days.**
* **Innovative and good marketing strategies lead bring more customers.**
* **Working on feedback and Improving the customer service.**

**C. For the States identified above, do the Root Cause analysis for their performance across a variety of metrics.**

**You can utilize the following metrics and explore a few yourself as well by analysing the data.**

**1 Category level Sales**

**Query:**

SELECT \* FROM (

SELECT customer\_state AS State,PRODUCT\_CATEGORY\_NAME , SUM(payment\_value)AS Sales

FROM olist\_orders\_dataset AS a

JOIN olist\_order\_payments\_dataset AS b

ON a.order\_id = b.order\_id

JOIN olist\_customers\_dataset AS c

ON a.customer\_id= c.customer\_id

JOIN olist\_order\_items\_dataset AS d

ON A.order\_id= D.order\_id

JOIN olist\_products\_dataset AS e

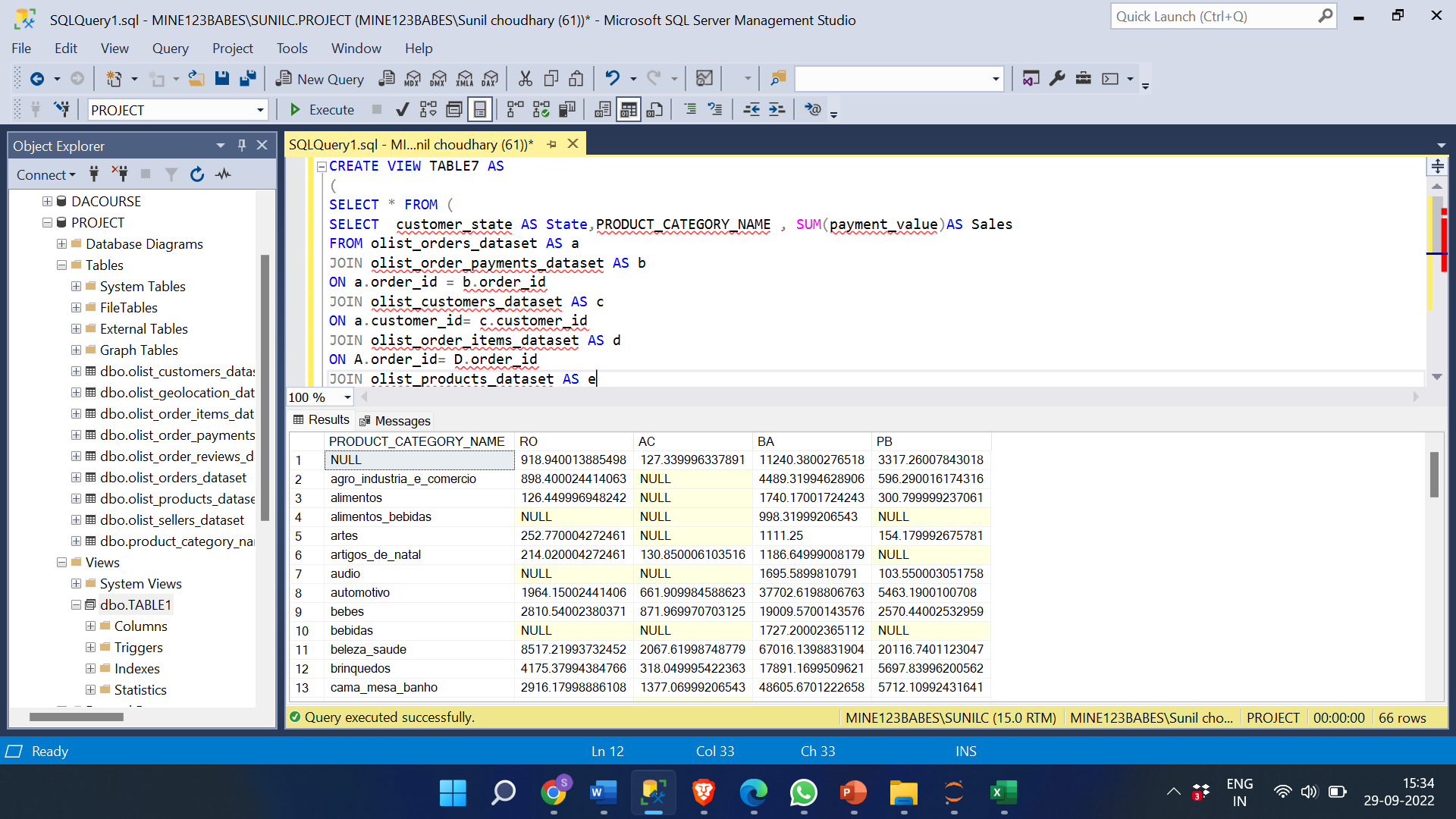
ON E.product\_id= D.product\_id

WHERE customer\_state IN ('RO', 'AC','BA', 'PB')

GROUP BY customer\_state ,PRODUCT\_CATEGORY\_NAME ) AS f

PIVOT

(MAX([Sales]) FOR [state] IN ([RO],[AC],[BA],[PB])) AS g



**2 Category level orders:**

**Query:**

SELECT \* FROM (

SELECT customer\_state AS state, PRODUCT\_CATEGORY\_NAME, COUNT(A.order\_id) AS No\_of\_orders FROM olist\_orders\_dataset a

JOIN olist\_customers\_dataset AS b

ON a.customer\_id= b.customer\_id

JOIN olist\_order\_items\_dataset AS c

ON A.order\_id= C.order\_id

JOIN olist\_products\_dataset AS d

ON D.product\_id= C.product\_id

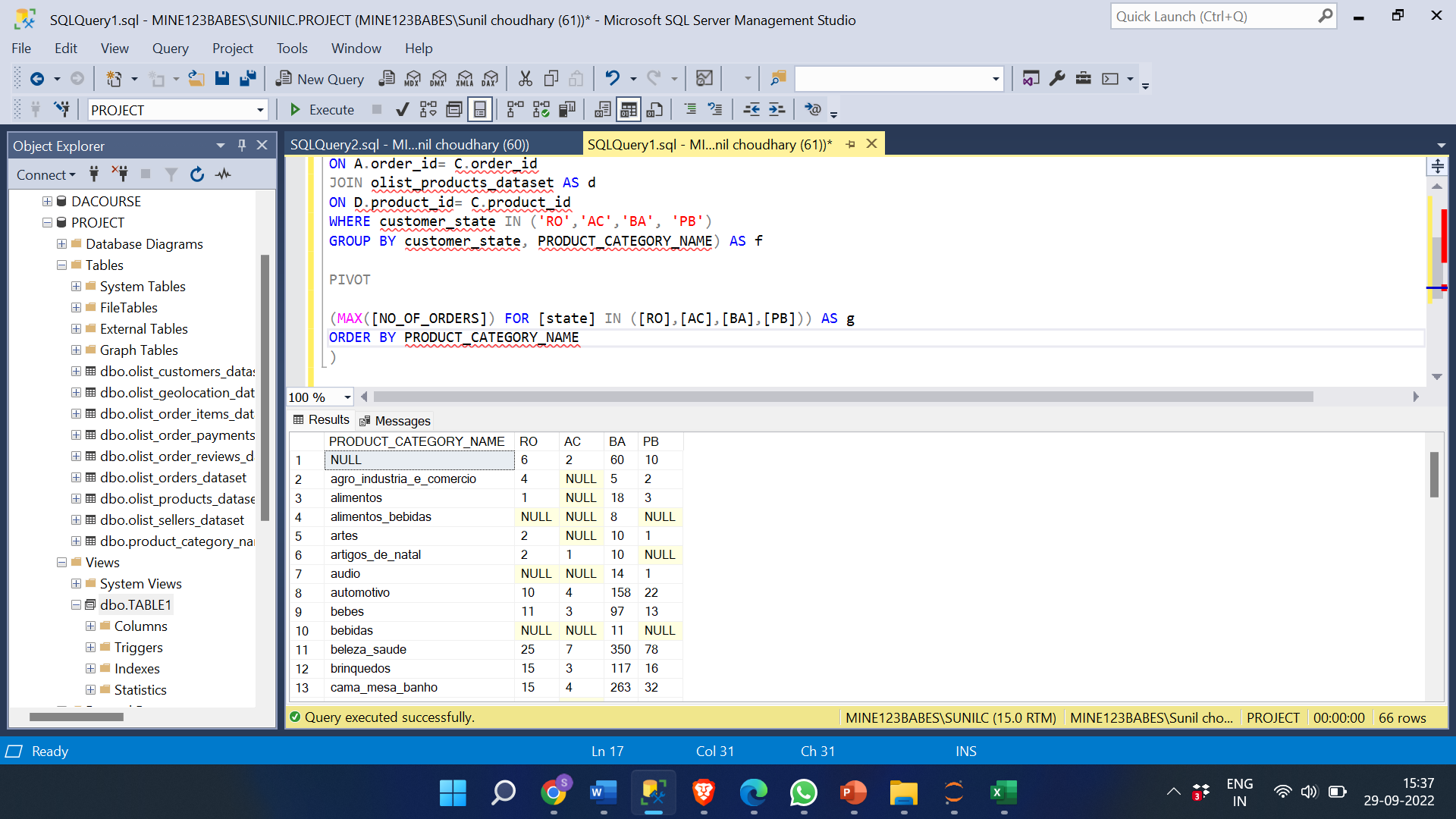
WHERE customer\_state IN ('RO','AC','BA', 'PB')

GROUP BY customer\_state, PRODUCT\_CATEGORY\_NAME) AS f

PIVOT

(MAX([NO\_OF\_ORDERS]) FOR [state] IN ([RO],[AC],[BA],[PB])) AS g

ORDER BY PRODUCT\_CATEGORY\_NAME



**INSIGHT:**

* **The most popular categories among all four states are beleza\_saude,** **esporte\_lazer,**

**cama\_mesa\_banho,** **informatica\_acessorios**, **moveis\_decoracao. There can be number of reasons for this such as demand, value, quality, need, identity, price, recommendations etc.**

* **least selling categories are: portateis\_cozinha\_e\_preparadores\_de\_alimentos,** **fraldas\_higiene,** **livros\_importados,** **fashion\_roupa\_infanto\_juvenil,** **dvds\_blu\_ray.**
* **The opposite can be said to the categories selling least products, reasons such as not demanded by many, over price, bad-quality, bad reviews etc.**

**1. Post-order reviews:**

SELECT customer\_state AS State, a.order\_id, review\_score, review\_comment\_message FROM olist\_order\_reviews\_dataset AS a

JOIN olist\_orders\_dataset AS b

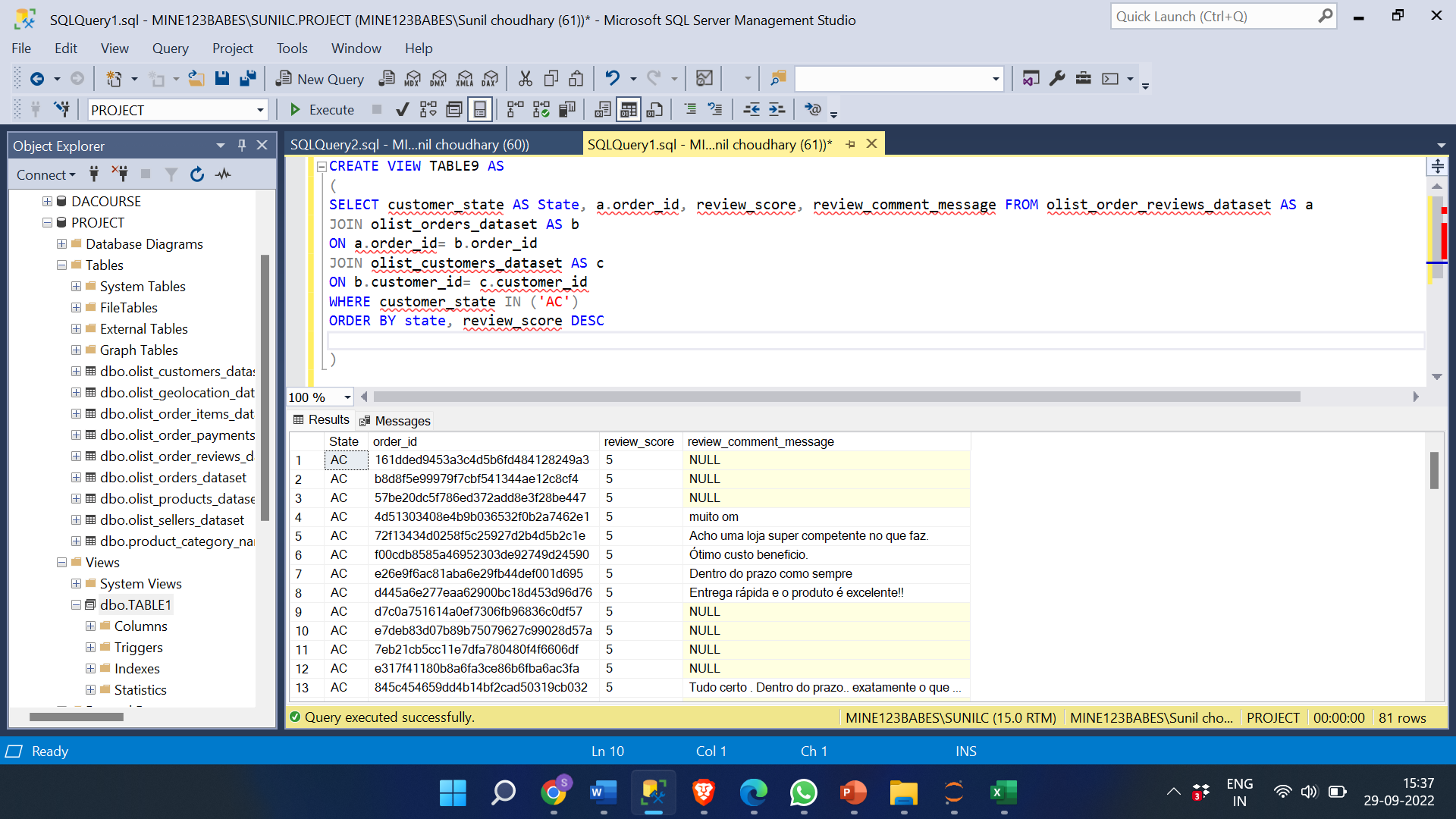
ON a.order\_id= b.order\_id

JOIN olist\_customers\_dataset AS c

ON b.customer\_id= c.customer\_id

WHERE customer\_state IN ('AC')

ORDER BY state, review\_score DESC



2. SELECT customer\_state AS State, a.order\_id, review\_score, review\_comment\_message FROM olist\_order\_reviews\_dataset AS a

JOIN olist\_orders\_dataset AS b

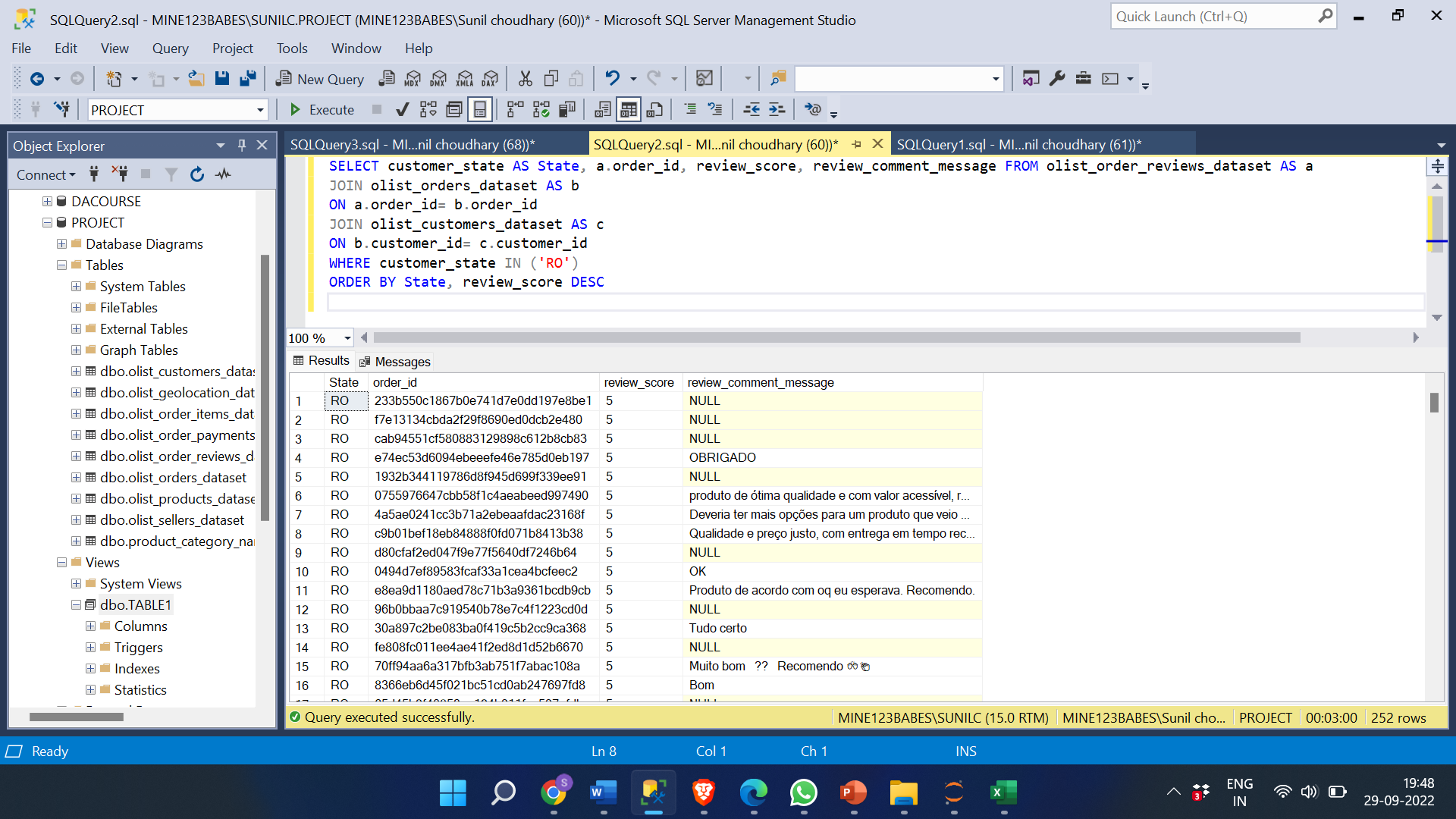
ON a.order\_id= b.order\_id

JOIN olist\_customers\_dataset AS c

ON b.customer\_id= c.customer\_id

WHERE customer\_state IN ('RO')

ORDER BY State, review\_score DESC



3. SELECT customer\_state AS State, a.order\_id, review\_score, review\_comment\_message FROM olist\_order\_reviews\_dataset AS a

JOIN olist\_orders\_dataset AS b

ON a.order\_id= b.order\_id

JOIN olist\_customers\_dataset AS c

ON b.customer\_id= c.customer\_id

WHERE customer\_state IN ('BA')

ORDER BY State, review\_score DESC



4. SELECT customer\_state AS state, a.order\_id, review\_score, review\_comment\_message FROM olist\_order\_reviews\_dataset AS a

JOIN olist\_orders\_dataset AS b

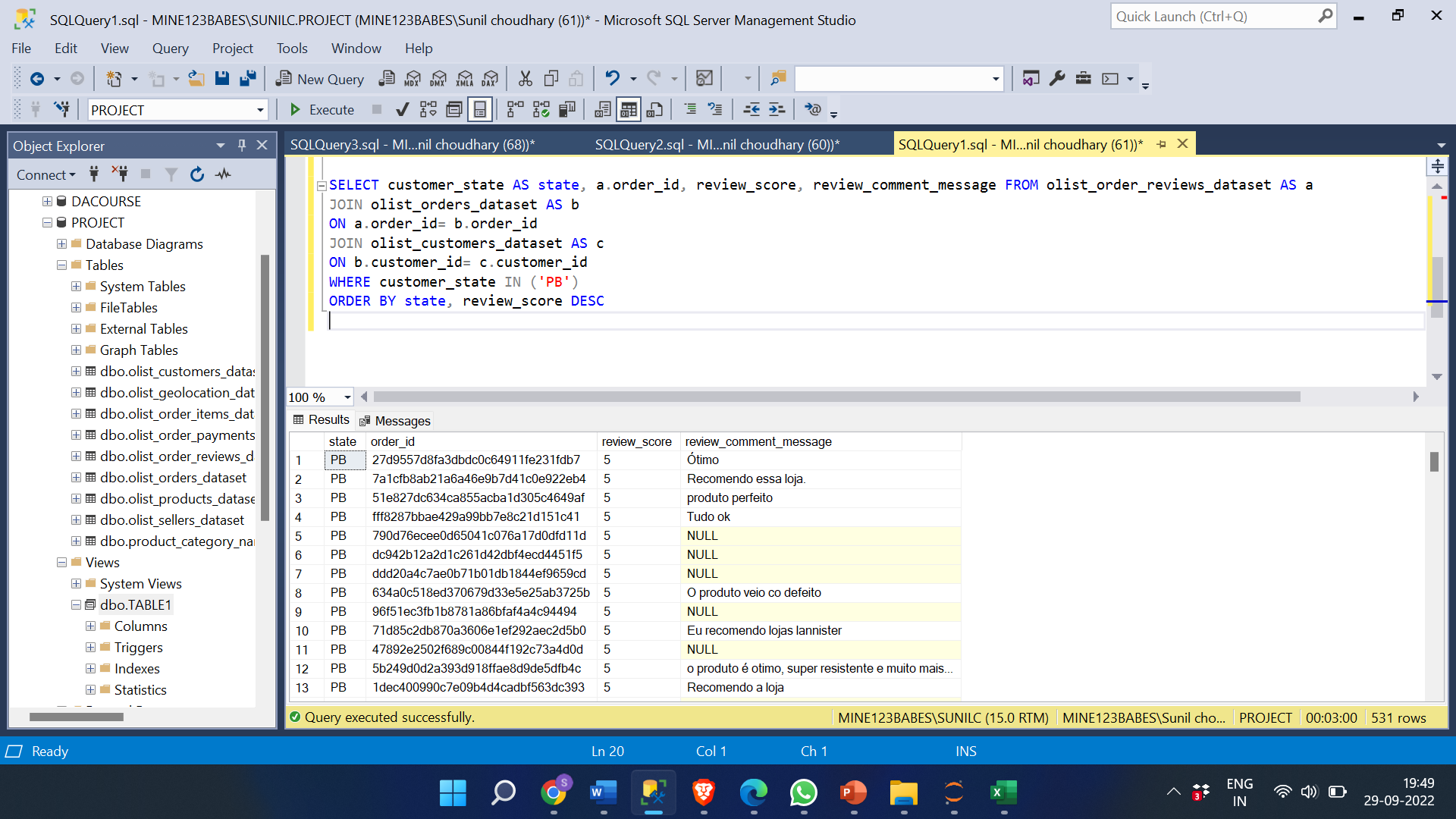
ON a.order\_id= b.order\_id

JOIN olist\_customers\_dataset AS c

ON b.customer\_id= c.customer\_id

WHERE customer\_state IN ('PB')

ORDER BY state, review\_score DESC



**INSIGHT:**

Reviews or feedback helps customer whether to buy that products or not. Postive Reviews helps to increase sales. Reviews for BA and PB are 4 and 5 for majority orders which helped in higher sales. Opposite for RO and AC.

**Seller performance in terms of deliveries.**

**1. AVERAGE DELIVERY DAYS FOR EACH SELLER IN EACH STATE.**

QUERY:

SELECT \* FROM (SELECT CUSTOMER\_STATE, SELLER\_ID,AVG( CONVERT(FLOAT,DATEDIFF(DAY,ORDER\_DELIVERED\_CARRIER\_DATE,ORDER\_DELIVERED\_CUSTOMER\_DATE))) AS AVG\_DELIVERY\_DAYS

FROM olist\_customers\_dataset AS a

JOIN olist\_orders\_dataset AS b

ON A.customer\_id= B.customer\_id

JOIN olist\_order\_items\_dataset AS c

ON B.order\_id=C.order\_id

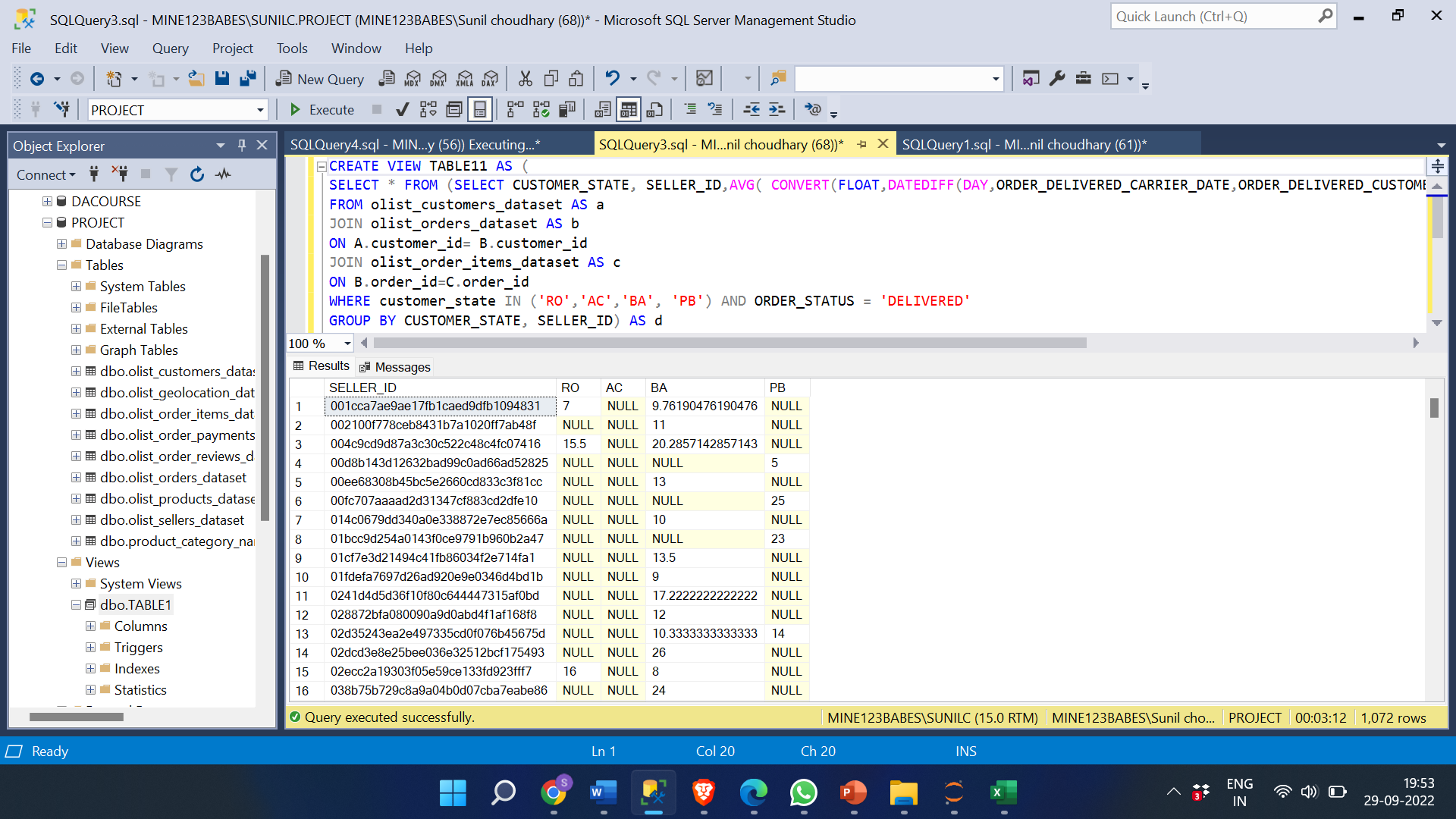
WHERE customer\_state IN ('RO','AC','BA', 'PB') AND ORDER\_STATUS = ‘DELIVERED'

GROUP BY CUSTOMER\_STATE, SELLER\_ID) AS d

PIVOT

(MAX([AVG\_DELIVERY\_DAYS]) FOR [CUSTOMER\_state] IN ([RO],[AC],[BA],[PB])) AS f

ORDER BY seller\_id



**INSIGHT:**

Different states have different sellers which deliver the order fastest as in most cases seller belongs to the same state as the customer. Some of the sellers deliver products to different states as well.

**2. NO OF DELIVERIES DONE BY EACH SELLER IN EACH STATE.**

SELECT \* FROM (

SELECT CUSTOMER\_STATE, SELLER\_ID,COUNT(C.ORDER\_ID) AS NO\_OF\_DELIVERIES

FROM olist\_customers\_dataset AS a

JOIN olist\_orders\_dataset AS b

ON A.customer\_id= B.customer\_id

JOIN olist\_order\_items\_dataset AS c

ON B.order\_id=C.order\_id

WHERE customer\_state IN ('RO','AC','BA', 'PB')

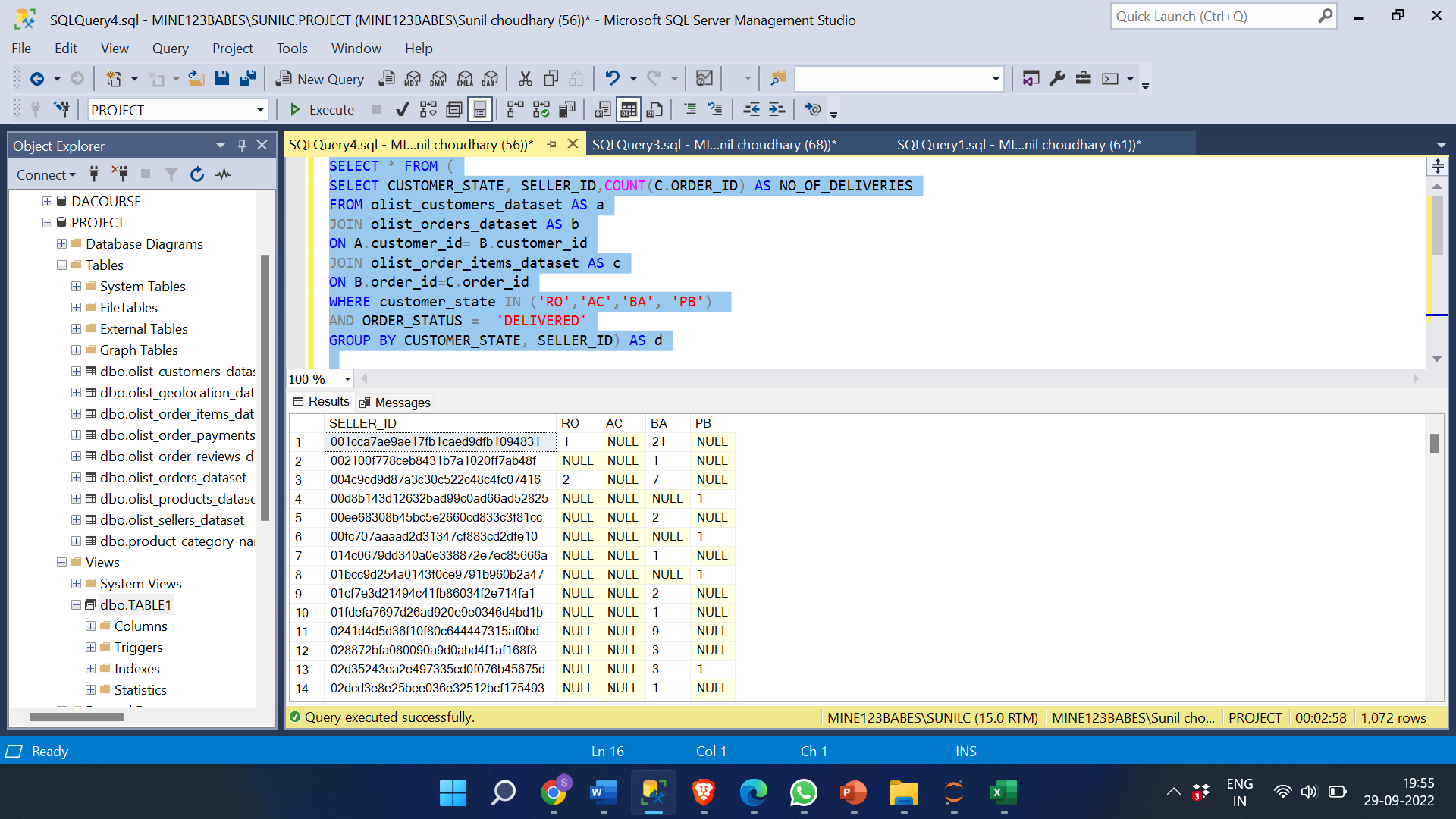
AND ORDER\_STATUS = 'DELIVERED'

GROUP BY CUSTOMER\_STATE, SELLER\_ID) AS d

PIVOT

(MAX([NO\_OF\_DELIVERIES]) FOR [CUSTOMER\_state] IN ([RO],[AC],[BA],[PB])) AS f

ORDER BY seller\_id



**INSIGHTS:**

The seller with most no of deliveries in State BA is seller\_id: 6560211a19b47992c3666cc44a7e94c0 with 91 successful deliveries.

The seller with most no of deliveries in State PB is seller\_id: cc419e0650a3c5ba77189a1882b7556a with 14 deliveries.

The seller with most no of deliveries in State RO is seller\_id: fa40cc5b934574b62717c68f3d678b6d with 9 deliveries.

The seller with most no of deliveries in State AC is seller\_id:

5a8e7d5003a1f221f9e1d6e411de7c23 with 6 deliveries.

**D) Product-level sales & orders placed**

**D1) PRODUCT LEVEL ORDERS**

**QUERY:**

SELECT \* FROM (

SELECT customer\_state AS state, C.product\_id, COUNT(A.order\_id) AS No\_of\_orders FROM olist\_orders\_dataset AS a

JOIN olist\_customers\_dataset AS b

ON a.customer\_id= b.customer\_id

JOIN olist\_order\_items\_dataset AS c

ON A.order\_id= C.order\_id

JOIN olist\_products\_dataset AS d

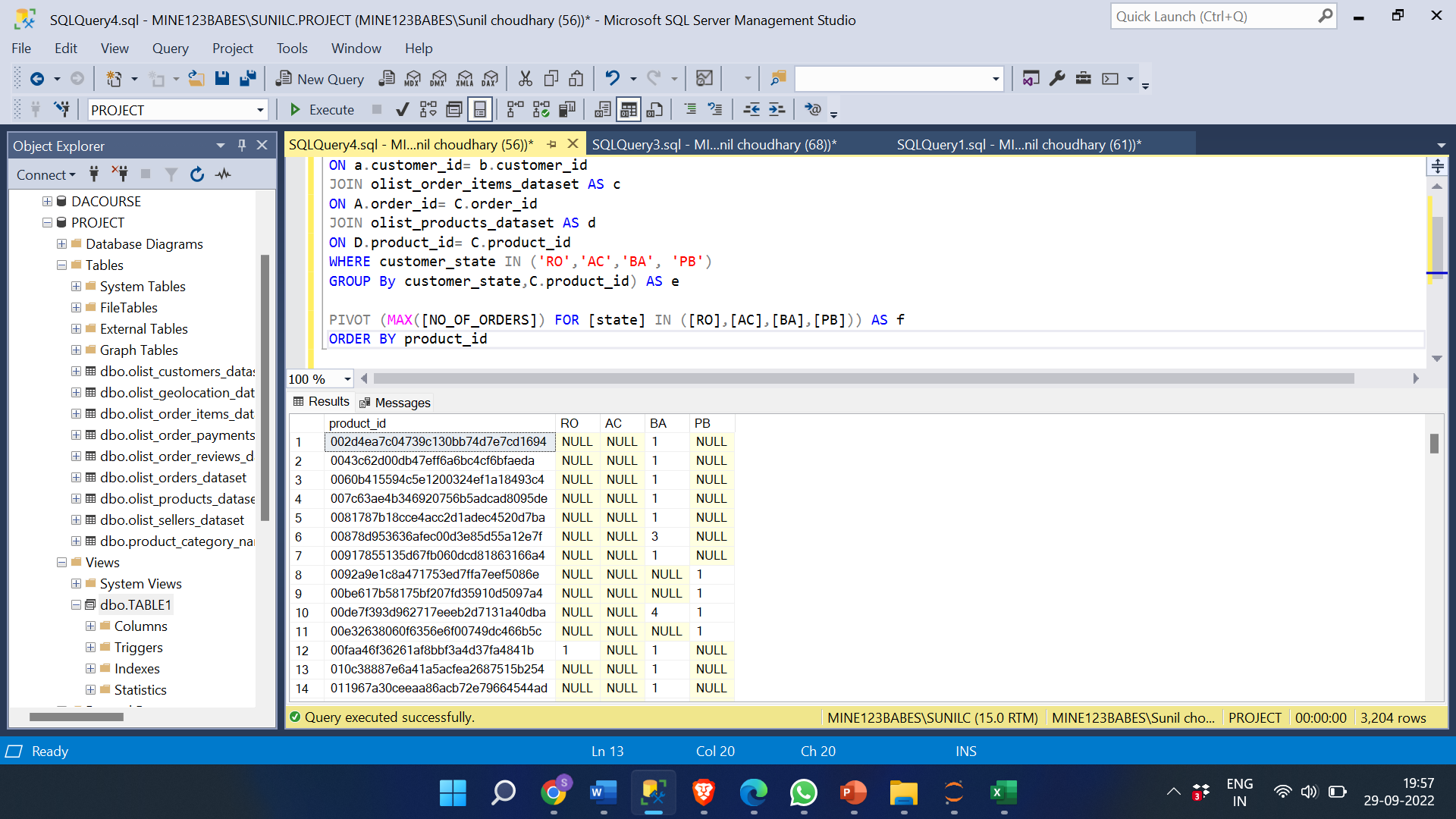
ON D.product\_id= C.product\_id

WHERE customer\_state IN ('RO','AC','BA', 'PB')

GROUP By customer\_state,C.product\_id) AS e

PIVOT (MAX([NO\_OF\_ORDERS]) FOR [state] IN ([RO],[AC],[BA],[PB])) AS f

ORDER BY product\_id



**1) PRODUCT LEVEL SALES**

QUERY:

SELECT \* FROM (

SELECT customer\_state AS State,D.product\_id , SUM(payment\_value)AS Sales

FROM olist\_orders\_dataset AS a

JOIN olist\_order\_payments\_dataset AS b

ON a.order\_id = b.order\_id

JOIN olist\_customers\_dataset AS c

ON a.customer\_id= c.customer\_id

JOIN olist\_order\_items\_dataset AS d

ON A.order\_id= D.order\_id

JOIN olist\_products\_dataset AS e

ON E.product\_id= D.product\_id

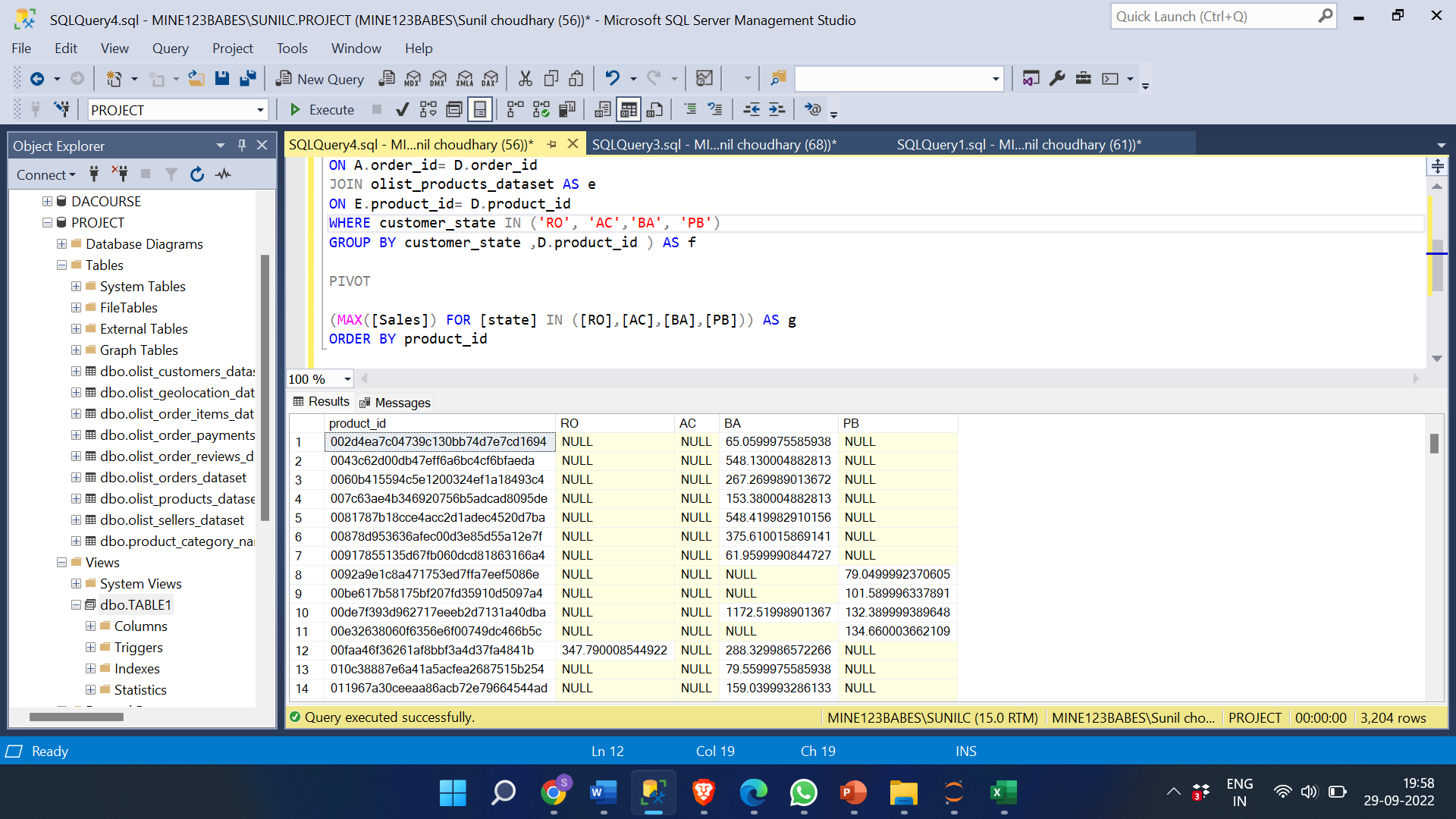
WHERE customer\_state IN ('RO', 'AC','BA', 'PB')

GROUP BY customer\_state ,D.product\_id ) AS f

PIVOT

(MAX([Sales]) FOR [state] IN ([RO],[AC],[BA],[PB])) AS g

ORDER BY product\_id



**INSIGHTS:**

**The most popular products are sold for similar reasons that the most popular categories are. It depends on the product's demand and need, usability, cost, quality, credibility, and recommendations.**

**Poor reviews, poor quality, insufficient marketing, an excessive price, low demand, and one more factor that may contribute to a product's low sales is the size of the item. People don't like to purchase big, bulky items online. These things have the potential to harm a product's brand.**

**A product's sales can be increased by effective marketing, fresh branding and packaging, message and sales tactics, discounts and attractive offers, and product redesigns that make them smaller and more streamlined.**

**E1) % of orders delivered earlier than the expected date.**

**QUERY:**

WITH CTE AS (

SELECT customer\_state AS State, COUNT(order\_id) AS TOTAL\_No\_of\_orders FROM olist\_orders\_dataset AS a

JOIN olist\_customers\_dataset AS b

ON a.customer\_id= b.customer\_id

WHERE customer\_state IN ('RO','AC','BA', 'PB') AND order\_status='DELIVERED'

GROUP BY customer\_state)

,

CTE1 AS (

SELECT customer\_state AS State, COUNT(order\_id) AS No\_of\_orders FROM olist\_orders\_dataset AS a

JOIN olist\_customers\_dataset AS b

ON a.customer\_id= b.customer\_id

WHERE customer\_state IN ('RO','AC','BA', 'PB') AND order\_delivered\_customer\_date < order\_estimated\_delivery\_date AND order\_status='DELIVERED'

GROUP BY customer\_state)

,

CTE2 AS (

SELECT CTE.STATE,NO\_OF\_ORDERS,TOTAL\_NO\_OF\_ORDERS

FROM CTE

JOIN CTE1

ON CTE.STATE= CTE1.STATE

)

SELECT \*, (CONVERT(FLOAT,NO\_OF\_ORDERS)/CONVERT(FLOAT,TOTAL\_NO\_OF\_ORDERS))\*100 AS PERCENT\_OF\_ORDERS\_DELIVERED\_EARLIER FROM CTE2

ORDER BY state



**E2) % of orders delivered Later than the expected date.**

**QUERY:**

WITH CTE AS (

SELECT customer\_state AS State, COUNT(order\_id) AS TOTAL\_No\_of\_orders FROM olist\_orders\_dataset AS a

JOIN olist\_customers\_dataset AS b

ON a.customer\_id= b.customer\_id

WHERE customer\_state IN ('RO','AC','BA', 'PB') AND order\_status='DELIVERED'

GROUP BY customer\_state)

,

CTE1 AS (

SELECT customer\_state AS State, COUNT(order\_id) AS No\_of\_orders FROM olist\_orders\_dataset AS a

JOIN olist\_customers\_dataset AS b

ON a.customer\_id= b.customer\_id

WHERE customer\_state IN ('RO','AC','BA', 'PB') AND order\_delivered\_customer\_date > order\_estimated\_delivery\_date AND order\_status='DELIVERED'

GROUP BY customer\_state)

,

CTE2 AS(

SELECT CTE.STATE,NO\_OF\_ORDERS,TOTAL\_NO\_OF\_ORDERS

FROM CTE

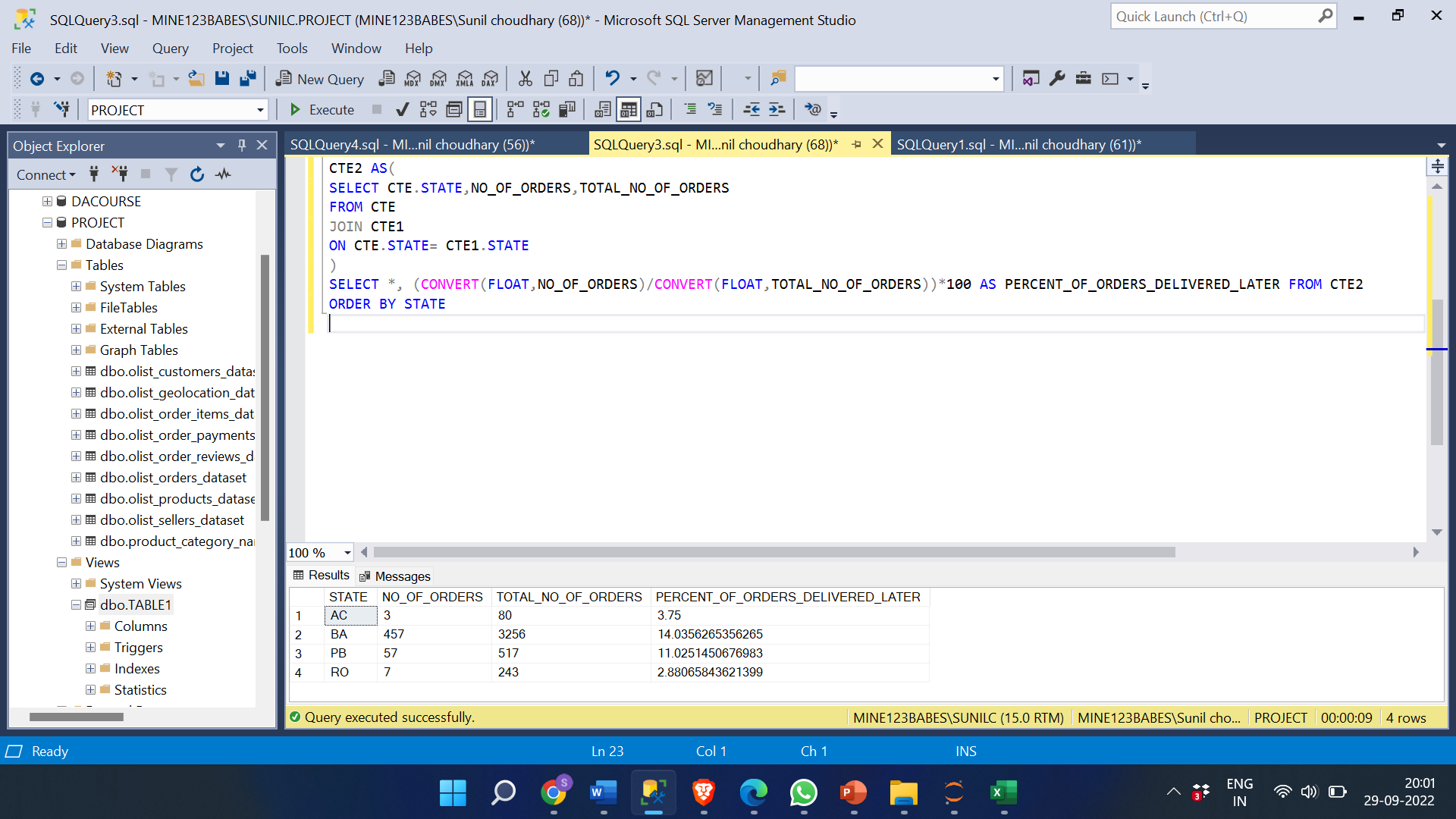
JOIN CTE1

ON CTE.STATE= CTE1.STATE

)

SELECT \*, (CONVERT(FLOAT,NO\_OF\_ORDERS)/CONVERT(FLOAT,TOTAL\_NO\_OF\_ORDERS))\*100 AS PERCENT\_OF\_ORDERS\_DELIVERED\_LATER FROM CTE2

ORDER BY STATE



**INSIGHT:**

**The most frequent reasons for delivery delays include issues with the vehicle, misplaced parcels, shipments coming from far away, bad weather, traffic, high volume shipments, customs, etc.**

**A product being delivered sooner than expected fosters confidence in both the seller and the business. This can result in more orders, more daily clients, and finally more sales.**

**The data above shows that the percentage of orders that are delivered later in states like PB and BA is larger than it is in RO and AC. There may be various explanations for this, but the fact that these states are in a metropolitan area with significant traffic volumes is probably the most important. Delivering the orders ahead of schedule can assist grow the client base. This can be accomplished by hiring more vendors and shippers in these states as well as by marketing. Also, to increase sales in rural areas more shippers should start sending products to these states this can lead to more customers using these platforms to buy products.**

**F) MOST COMMON PAYMENT METHOD AMONG ALL THESE STATES.**

SELECT \* FROM (

SELECT customer\_state AS State, payment\_type, COUNT(b.order\_id) AS No\_of\_orders from olist\_order\_payments\_dataset AS a

JOIN olist\_orders\_dataset AS b

ON a.order\_id= b.order\_id

JOIN olist\_customers\_dataset AS c

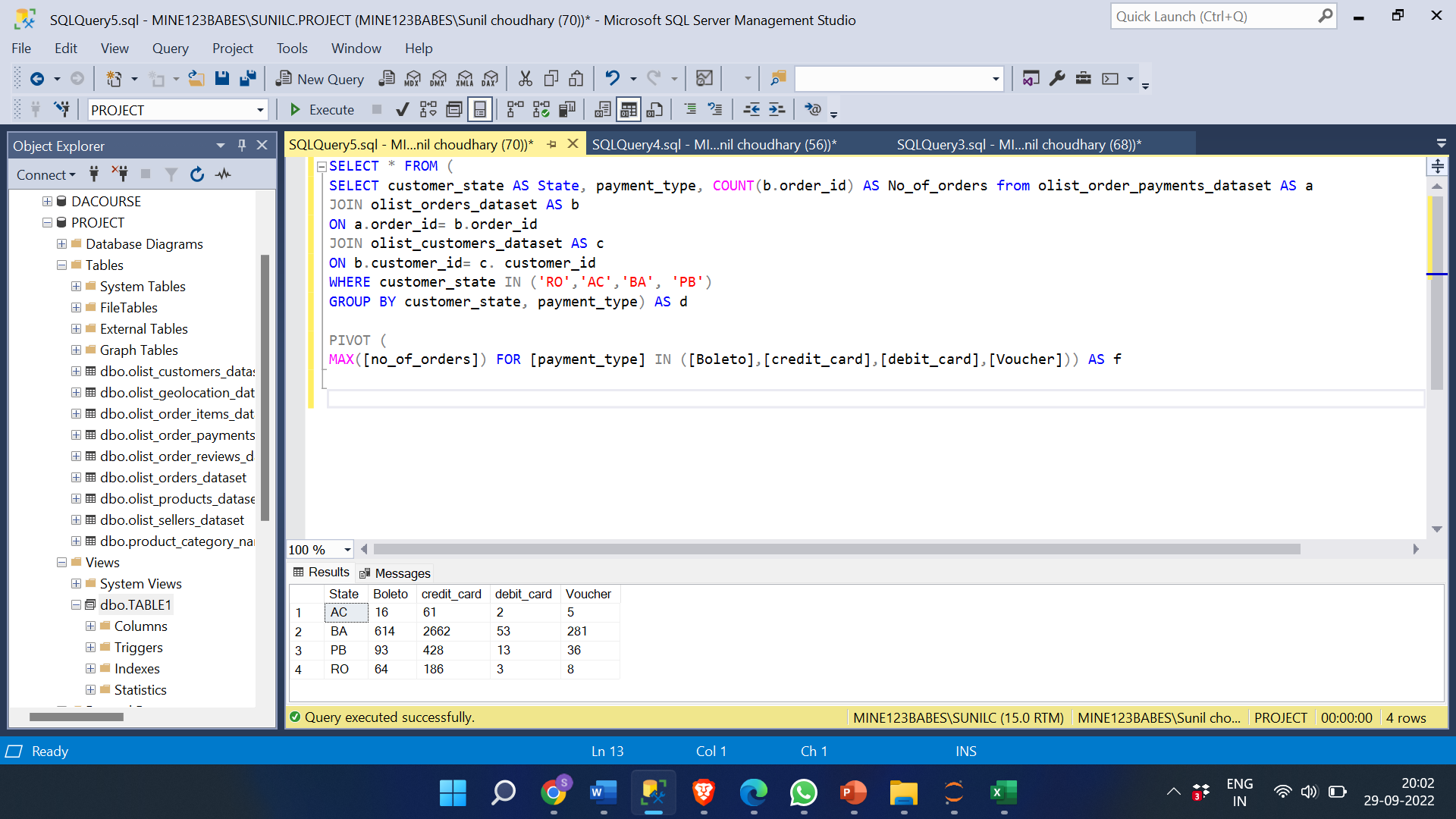
ON b.customer\_id= c. customer\_id

WHERE customer\_state IN ('RO','AC','BA', 'PB')

GROUP BY customer\_state, payment\_type) AS d

PIVOT (

MAX([no\_of\_orders]) FOR [payment\_type] IN ([Boleto],[credit\_card],[debit\_card],[Voucher])) AS f



**INSIGHT:**

**Credit cards are the most popular mode of payment across all states, and individuals in BA and PB also like to order using Boleto and vouchers.**

**Sales can be raised, and new clients can join by pushing the alternative mode of payment into the RO and AC states. Additionally, offering occasional discounts on the products can aid in attracting new clients and boosting sales.**