**Target BusinessCase Solution**

**Name: Arjun Shrinivas**

**Batch: July2022Beginner**

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

SELECT

table\_name

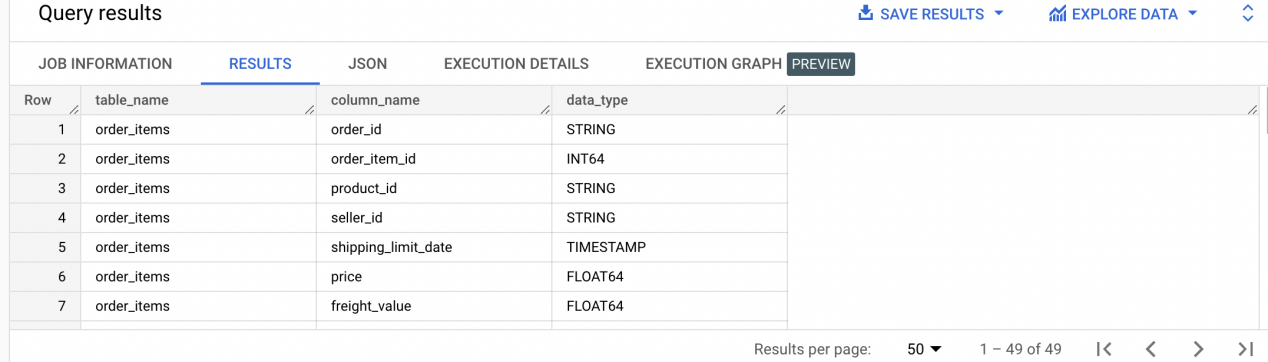
,column\_name

,data\_type

FROM

`BC01\_TargetDataSet.INFORMATION\_SCHEMA.COLUMNS`

;



1. **Time period for which the data is given**

SELECT

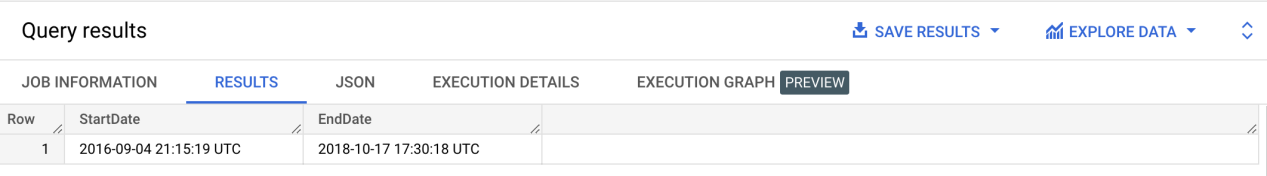
MIN(order\_purchase\_timestamp) AS StartDate

,MAX(order\_purchase\_timestamp) AS EndDate

FROM

`BC01\_TargetDataSet.orders`

;



1. **Cities and States of customers ordered during the given period**

SELECT DISTINCT

C.customer\_state

,C.customer\_city

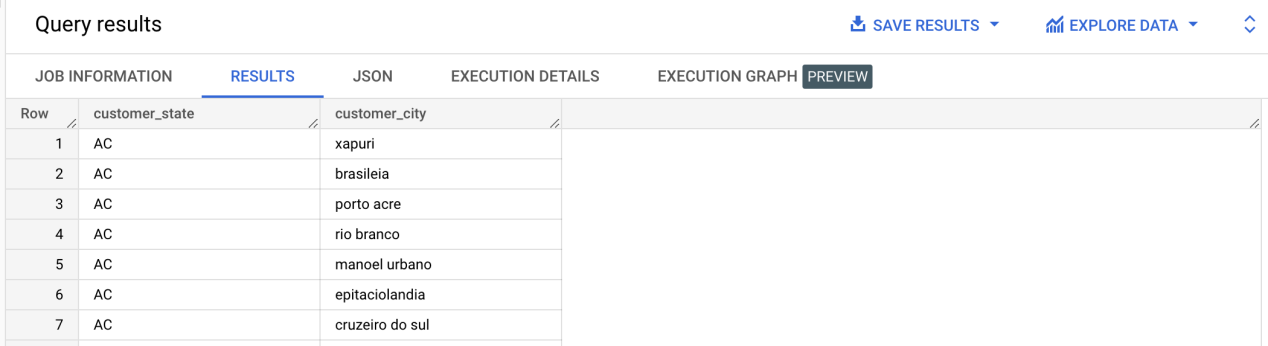
FROM

`BC01\_TargetDataSet.customers` C

INNER JOIN `BC01\_TargetDataSet.orders` O ON C.customer\_id = O.customer\_id

ORDER BY C.customer\_state

;



1. **In-depth Exploration:**
2. **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?**

WITH Temp1 AS (

SELECT

order\_id

,customer\_id

,DATE(order\_purchase\_timestamp) DateDetail

FROM `BC01\_TargetDataSet.orders`

),

Temp2 AS (

SELECT

COUNT(T1.order\_id) Counter

,EXTRACT(YEAR FROM T1.DateDetail) Year

,EXTRACT(MONTH FROM T1.DateDetail) Month

FROM Temp1 T1

GROUP BY EXTRACT(YEAR FROM T1.DateDetail),EXTRACT(MONTH FROM T1.DateDetail)

)

SELECT

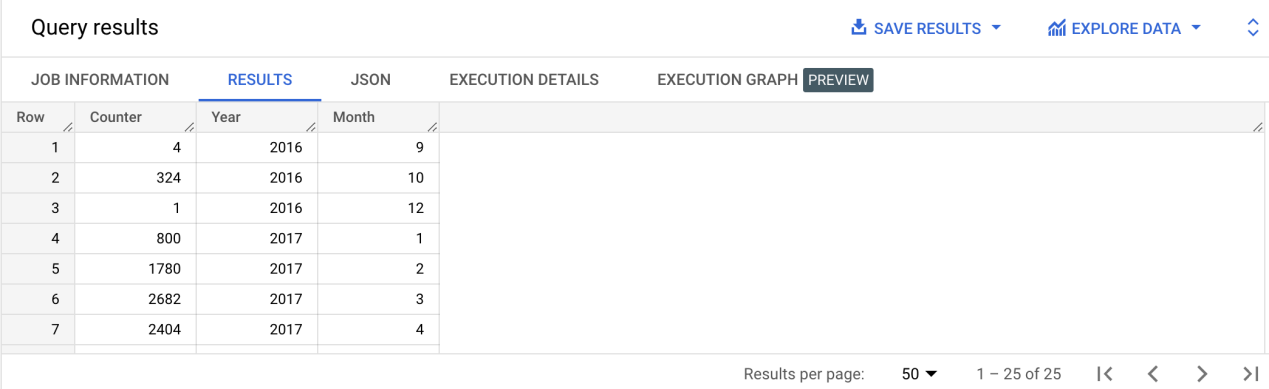
\*

FROM

Temp2 T2

ORDER BY T2.YEAR,T2.Month

;



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

WITH Temp1 AS (

SELECT

order\_id

,customer\_id

,EXTRACT(TIME FROM order\_purchase\_timestamp) TimeDetail

FROM `BC01\_TargetDataSet.orders`

),

Temp2 AS (

SELECT

T1.order\_id

,T1.customer\_id

,CASE

WHEN T1.TimeDetail BETWEEN '00:00:00' AND '05:59:00'

THEN 'Dusk'

WHEN T1.TimeDetail BETWEEN '06:00:00' AND '11:59:00'

THEN 'Morning'

WHEN T1.TimeDetail BETWEEN '12:00:00' AND '17:59:00'

THEN 'Afternoon'

WHEN T1.TimeDetail BETWEEN '18:00:00' AND '23:59:00'

THEN 'Night'

END AS TimeSlot

FROM

Temp1 T1

)

SELECT DISTINCT

COUNT(order\_id) OVER(PARTITION BY TimeSlot) CountByTime

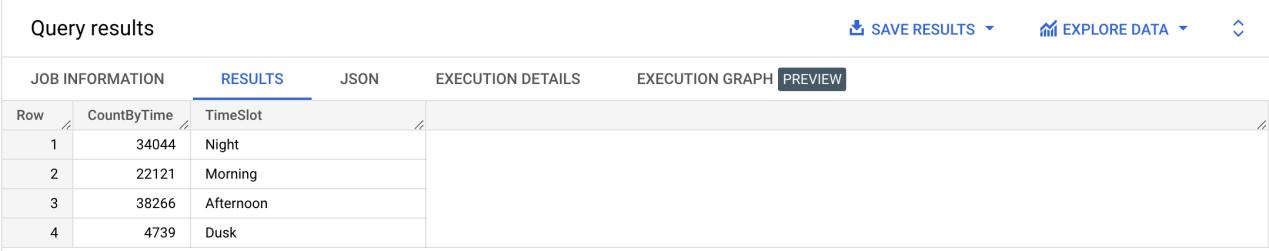
,TimeSlot

FROM

Temp2

WHERE TimeSlot IS NOT NULL

;



1. **Evolution of E-commerce orders in the Brazil region:**
2. **Get month on month orders by states**

WITH Temp1 AS (

SELECT

O.order\_id

,O.customer\_id

,DATE(O.order\_purchase\_timestamp) DateDetail

,C.customer\_state

FROM `BC01\_TargetDataSet.orders` O

INNER JOIN `BC01\_TargetDataSet.customers` C ON O.customer\_id = C.customer\_id

),

Temp2 AS (

SELECT

COUNT(T1.order\_id) CountByMonthByState

,EXTRACT(YEAR FROM T1.DateDetail) Year

,EXTRACT(MONTH FROM T1.DateDetail) Month

,T1.customer\_state

FROM Temp1 T1

GROUP BY EXTRACT(YEAR FROM T1.DateDetail),EXTRACT(MONTH FROM T1.DateDetail),T1.customer\_state

)

SELECT

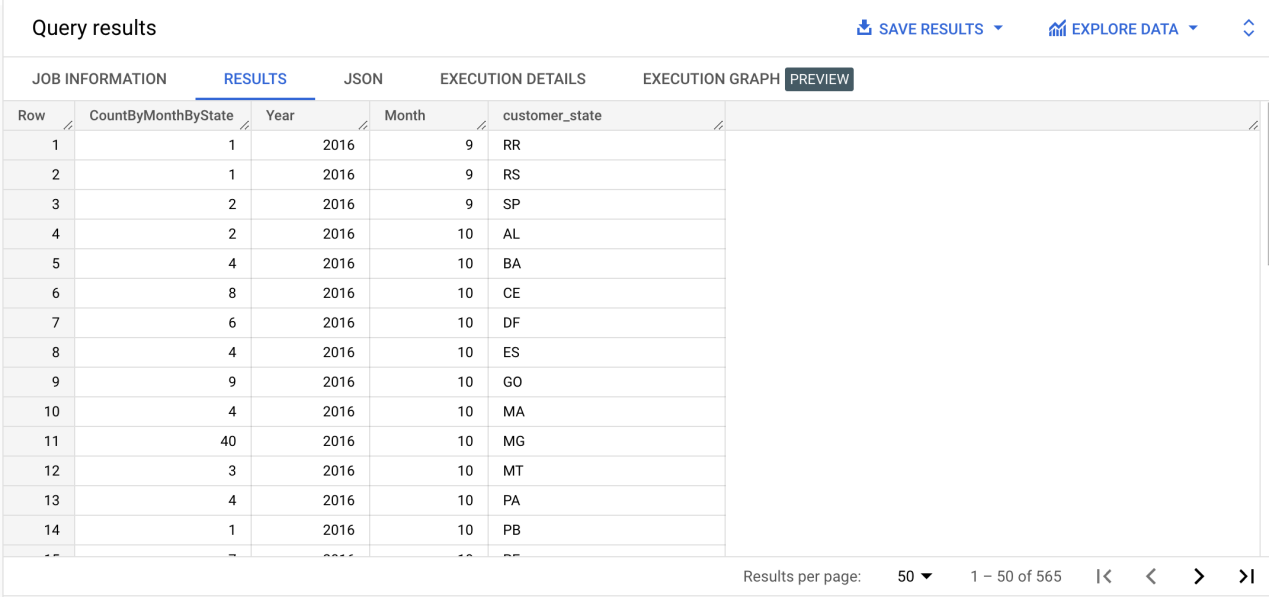
\*

FROM

Temp2 T2

ORDER BY T2.Year,T2.Month,T2.customer\_state

;



1. **Distribution of customers across the states in Brazil**

SELECT DISTINCT

customer\_state

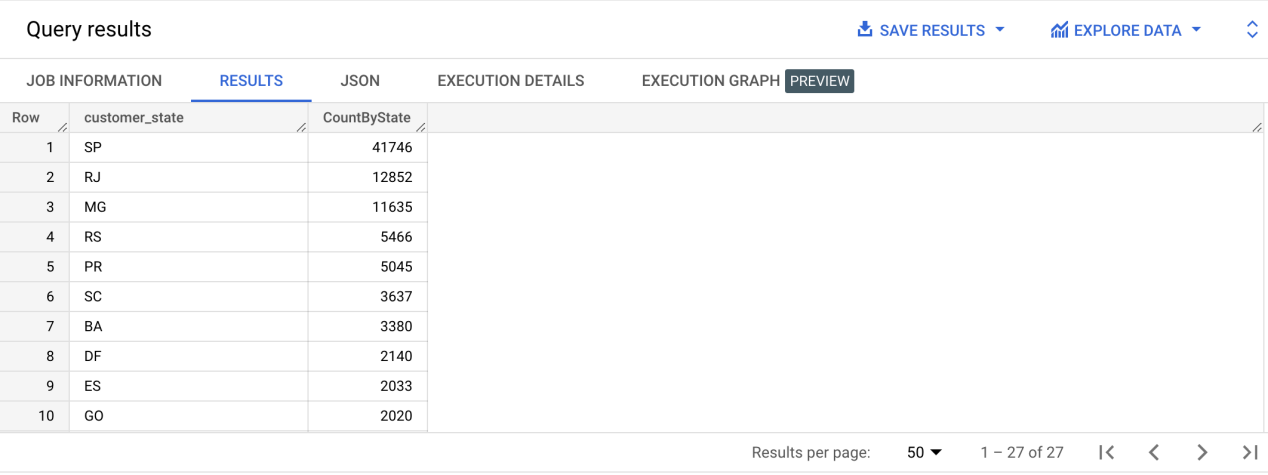
,COUNT(customer\_id) OVER(PARTITION BY customer\_state) CountByState

FROM

`BC01\_TargetDataSet.customers`

ORDER BY CountByState DESC

;



1. **Impact on Economy: Analyze the money movement by e-commerce by looking at order prices, freight and others.**
2. **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**

WITH Temp1 AS (

SELECT

P.order\_id

,P.payment\_value

,EXTRACT(MONTH FROM O.order\_purchase\_timestamp) Month

,EXTRACT(YEAR FROM O.order\_purchase\_timestamp) Year

FROM

`BC01\_TargetDataSet.payments` P

INNER JOIN `BC01\_TargetDataSet.orders` O ON P.order\_id = O.order\_id

ORDER BY Year,Month

),

Temp2 AS (

SELECT DISTINCT

Year

,Month

,ROUND(SUM(payment\_value) OVER(PARTITION BY Year,Month ORDER BY Year,Month),2) CurrentMonthSale

FROM Temp1

),

Temp3 AS (

SELECT

Year

,Month

,CurrentMonthSale

,LAG(CurrentMonthSale,1) OVER(ORDER BY Year,Month) PreviousMonthSale

,ROUND(CurrentMonthSale - LAG(CurrentMonthSale,1) OVER(ORDER BY Year,Month),2) AbsChange

FROM

Temp2

ORDER BY Year,Month

)

SELECT

Year

,Month

,CurrentMonthSale

,PreviousMonthSale

,AbsChange

,ROUND(((AbsChange / PreviousMonthSale) \* 100),2) PercentageChange

FROM

Temp3

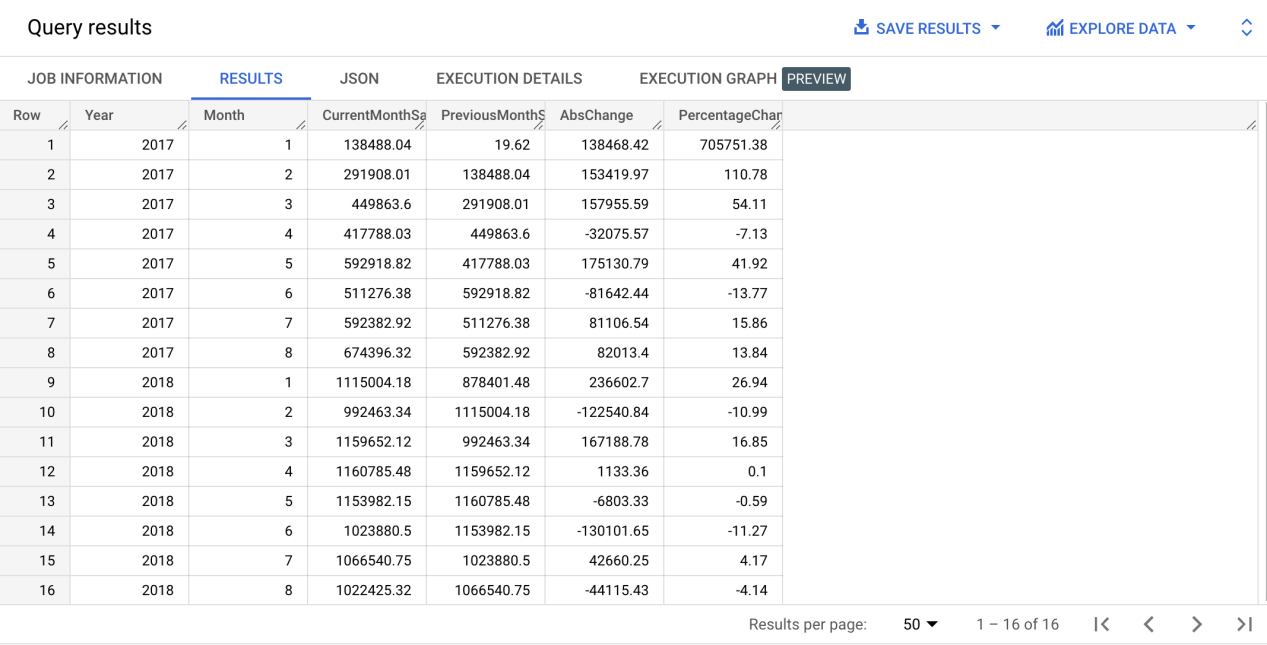
WHERE

(Year BETWEEN 2017 AND 2018)

AND

(Month BETWEEN 1 AND 8)

;



1. **Mean & Sum of price and freight value by customer state**

WITH Temp1 AS (

SELECT DISTINCT

ROUND(SUM(OI.price) OVER(PARTITION BY C.customer\_state),2) TotalPriceByState

,ROUND(AVG(OI.price) OVER(PARTITION BY C.customer\_state),2) AvgPriceByState

,ROUND(SUM(OI.freight\_value) OVER(PARTITION BY C.customer\_state),2) TotalFreightByState

,ROUND(AVG(OI.freight\_value) OVER(PARTITION BY C.customer\_state),2) AvgFreightByState

,C.customer\_state

FROM

`BC01\_TargetDataSet.order\_items` OI

INNER JOIN `BC01\_TargetDataSet.orders` O on OI.order\_id = O.order\_id

INNER JOIN `BC01\_TargetDataSet.customers` C on O.customer\_id = C.customer\_id

ORDER BY C.customer\_state

)

SELECT

TotalPriceByState

,AvgPriceByState

,TotalFreightByState

,AvgFreightByState

,ROUND((TotalPriceByState + TotalFreightByState),2) TotalCostByState

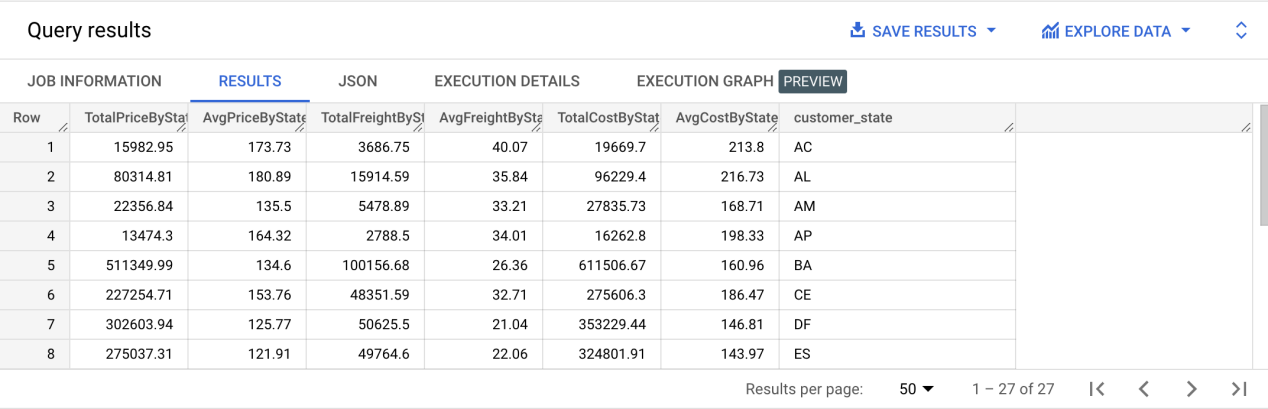
,ROUND((AvgPriceByState + AvgFreightByState),2) AvgCostByState

,customer\_state

FROM

Temp1

;



1. **Analysis on sales, freight and delivery time**
2. **Calculate days between purchasing, delivering and estimated delivery**

SELECT DISTINCT

order\_id

,DATE\_DIFF(order\_delivered\_customer\_date,order\_purchase\_timestamp,DAY) DaysToDeliver

,DATE\_DIFF(order\_estimated\_delivery\_date,order\_purchase\_timestamp,DAY) DaysEstimated

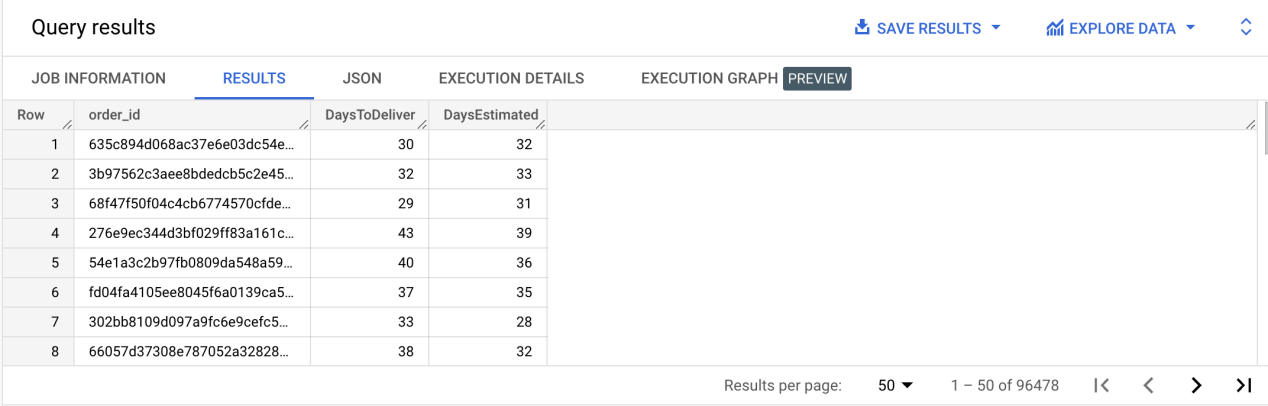
FROM

`BC01\_TargetDataSet.orders`

WHERE

order\_status = 'delivered'

;



1. **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 DISTINCT

order\_id

,DATE\_DIFF(order\_delivered\_customer\_date,order\_purchase\_timestamp,DAY) time\_to\_delivery

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

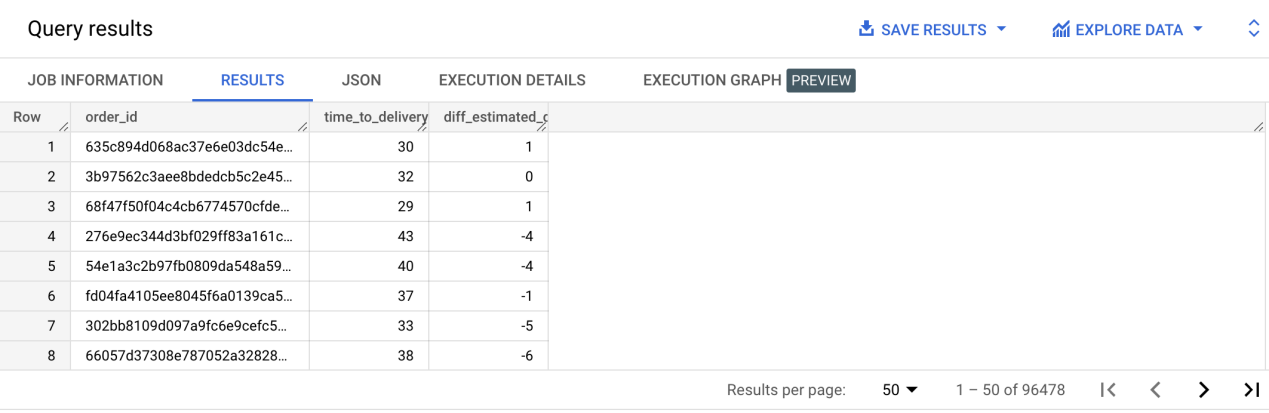
FROM

`BC01\_TargetDataSet.orders`

WHERE

order\_status = 'delivered'

;



1. **Group data by state, take mean of freight\_value, time\_to\_delivery, diff\_estimated\_delivery**

WITH Temp1 AS (

SELECT DISTINCT

OI.freight\_value

,C.customer\_state

,DATE\_DIFF(O.order\_delivered\_customer\_date,O.order\_purchase\_timestamp,DAY) time\_to\_delivery

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

FROM

`BC01\_TargetDataSet.orders` O

INNER JOIN `BC01\_TargetDataSet.order\_items` OI ON O.order\_id = OI.order\_id

INNER JOIN `BC01\_TargetDataSet.customers` C ON O.customer\_id = C.customer\_id

WHERE

order\_status = 'delivered'

)

SELECT DISTINCT

customer\_state

,ROUND(AVG(freight\_value) OVER(PARTITION BY customer\_state),2) mean\_freight\_value\_by\_state

,ROUND(AVG(time\_to\_delivery) OVER(PARTITION BY customer\_state),2) mean\_time\_to\_delivery\_by\_state

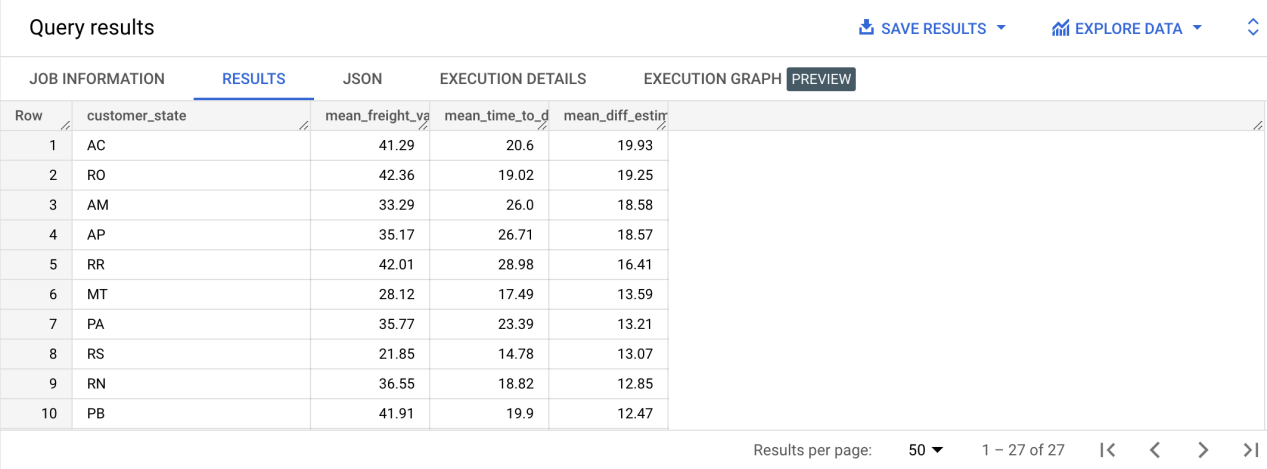
,ROUND(AVG(diff\_estimated\_delivery) OVER(PARTITION BY customer\_state),2) mean\_diff\_estimated\_delivery\_by\_state

FROM

Temp1

ORDER BY mean\_diff\_estimated\_delivery\_by\_state

;



1. **Sort the data to get the following:**

CREATE VIEW `BC01\_TargetDataSet.FreightAndDeliveryInfo` AS (

SELECT DISTINCT

OI.freight\_value

,C.customer\_state

,DATE\_DIFF(O.order\_delivered\_customer\_date,O.order\_purchase\_timestamp,DAY) time\_to\_delivery

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

FROM

`BC01\_TargetDataSet.orders` O

INNER JOIN `BC01\_TargetDataSet.order\_items` OI ON O.order\_id = OI.order\_id

INNER JOIN `BC01\_TargetDataSet.customers` C ON O.customer\_id = C.customer\_id

WHERE

order\_status = 'delivered'

);

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

SELECT DISTINCT

customer\_state

,ROUND(AVG(freight\_value) OVER(PARTITION BY customer\_state),2) mean\_freight\_value\_by\_state

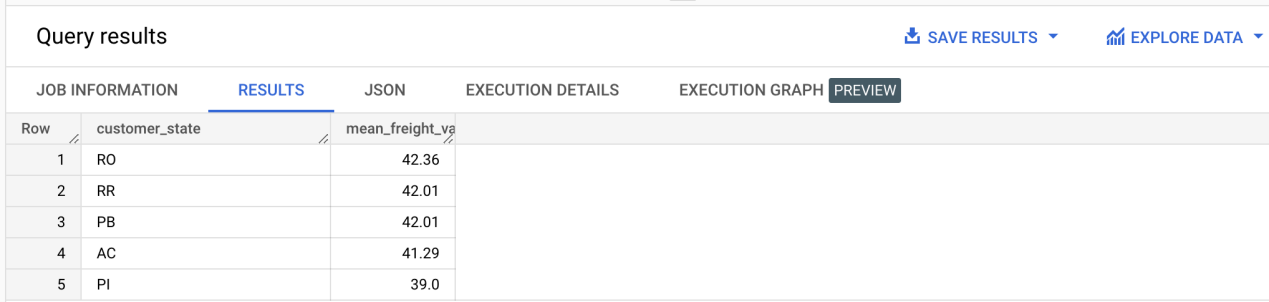
FROM

`BC01\_TargetDataSet.FreightAndDeliveryInfo`

ORDER BY mean\_freight\_value\_by\_state DESC

LIMIT 5

;



* + 1. **Top 5 states with lowest average freight value**

SELECT DISTINCT

customer\_state

,ROUND(AVG(freight\_value) OVER(PARTITION BY customer\_state),2) mean\_freight\_value\_by\_state

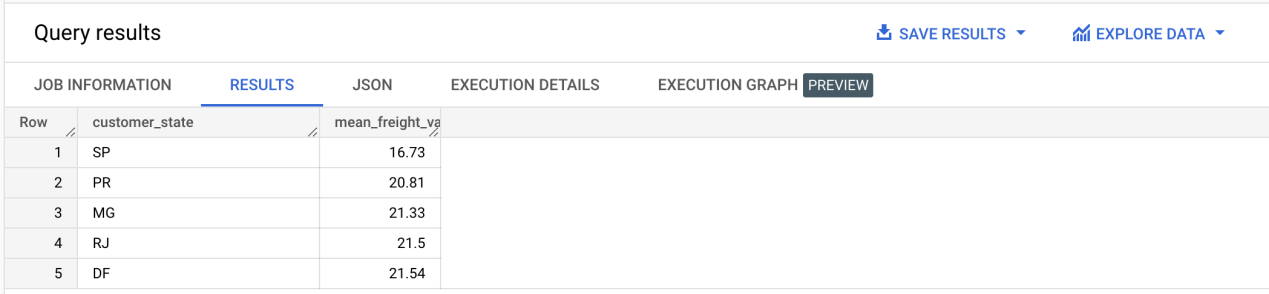
FROM

`BC01\_TargetDataSet.FreightAndDeliveryInfo`

ORDER BY mean\_freight\_value\_by\_state ASC

LIMIT 5

;



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

SELECT DISTINCT

customer\_state

,ROUND(AVG(time\_to\_delivery) OVER(PARTITION BY customer\_state),2) mean\_time\_to\_delivery\_by\_state

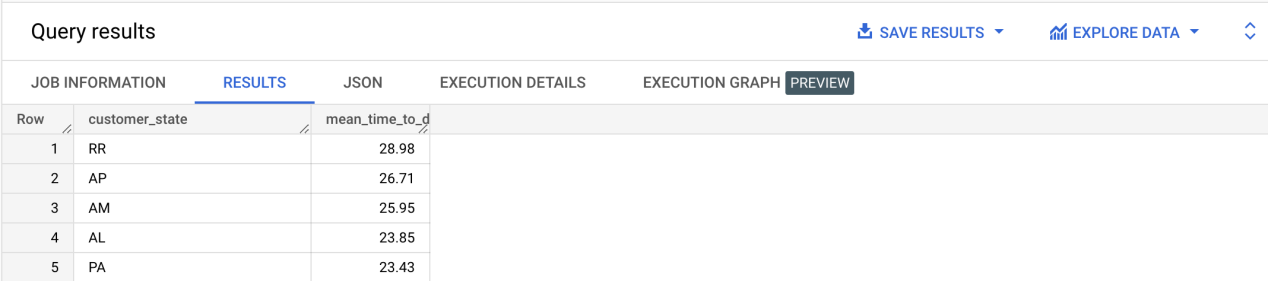
FROM

`BC01\_TargetDataSet.FreightAndDeliveryInfo`

ORDER BY mean\_time\_to\_delivery\_by\_state DESC

LIMIT 5

;



* + 1. **Top 5 states with lowest average time to delivery**

SELECT DISTINCT

customer\_state

,ROUND(AVG(time\_to\_delivery) OVER(PARTITION BY customer\_state),2) mean\_time\_to\_delivery\_by\_state

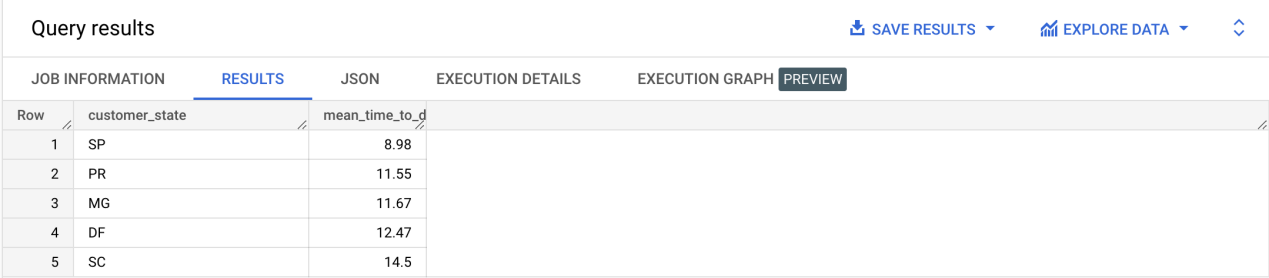
FROM

`BC01\_TargetDataSet.FreightAndDeliveryInfo`

ORDER BY mean\_time\_to\_delivery\_by\_state ASC

LIMIT 5

;



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

SELECT DISTINCT

customer\_state

,ROUND(AVG(diff\_estimated\_delivery) OVER(PARTITION BY customer\_state),2) mean\_diff\_estimated\_delivery\_by\_state

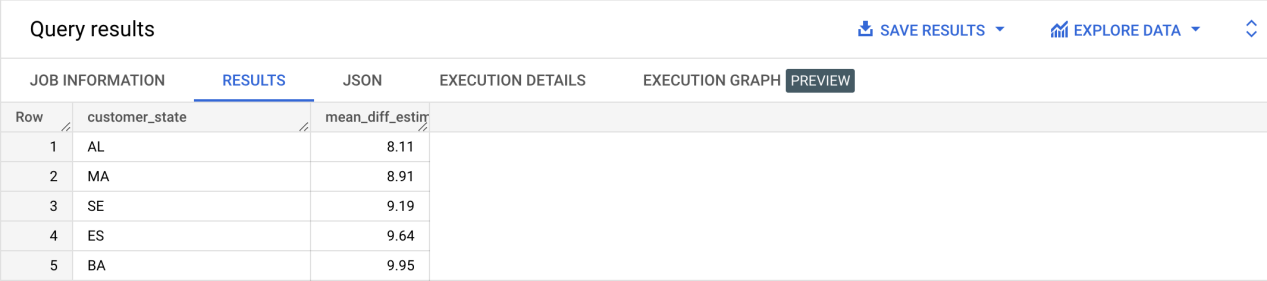
FROM

`BC01\_TargetDataSet.FreightAndDeliveryInfo`

ORDER BY mean\_diff\_estimated\_delivery\_by\_state ASC

LIMIT 5

;



* + 1. **Top 5 states where delivery is not so fast compared to estimated date**

SELECT DISTINCT

customer\_state

,ROUND(AVG(diff\_estimated\_delivery) OVER(PARTITION BY customer\_state),2) mean\_diff\_estimated\_delivery\_by\_state

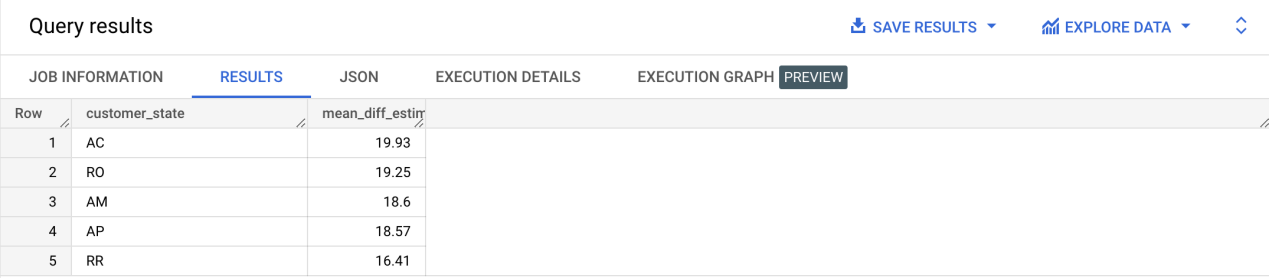
FROM

`BC01\_TargetDataSet.FreightAndDeliveryInfo`

ORDER BY mean\_diff\_estimated\_delivery\_by\_state DESC

LIMIT 5

;



1. **Payment type analysis:**

CREATE VIEW `BC01\_TargetDataSet.PaymentAnalysis` AS (

SELECT DISTINCT

P.order\_id

,P.payment\_type

,P.payment\_installments

,EXTRACT(YEAR FROM O.order\_purchase\_timestamp) order\_year

,EXTRACT(MONTH FROM O.order\_purchase\_timestamp) order\_month

FROM

`BC01\_TargetDataSet.payments` P

INNER JOIN `BC01\_TargetDataSet.orders` O ON P.order\_id = O.order\_id

ORDER BY order\_year,order\_month

);

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

SELECT DISTINCT

COUNT(order\_id) OVER(PARTITION BY payment\_type,order\_year,order\_month) order\_count

,payment\_type

,order\_year

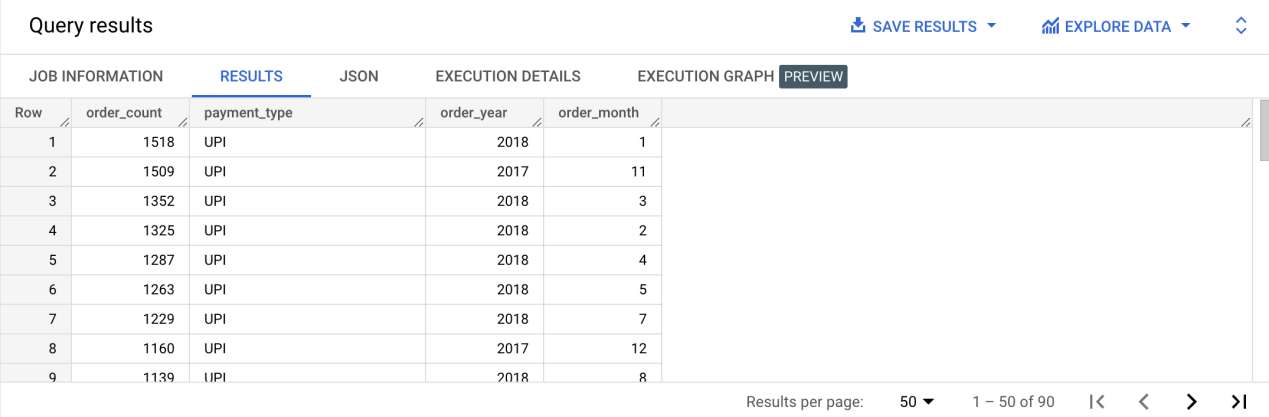
,order\_month

FROM

`BC01\_TargetDataSet.PaymentAnalysis`

ORDER BY payment\_type ASC,order\_count DESC

;



1. **Count of orders based on the no. of payment installments**

SELECT DISTINCT

COUNT(order\_id) OVER(PARTITION BY payment\_installments,order\_year,order\_month) order\_count

,payment\_installments

,order\_year

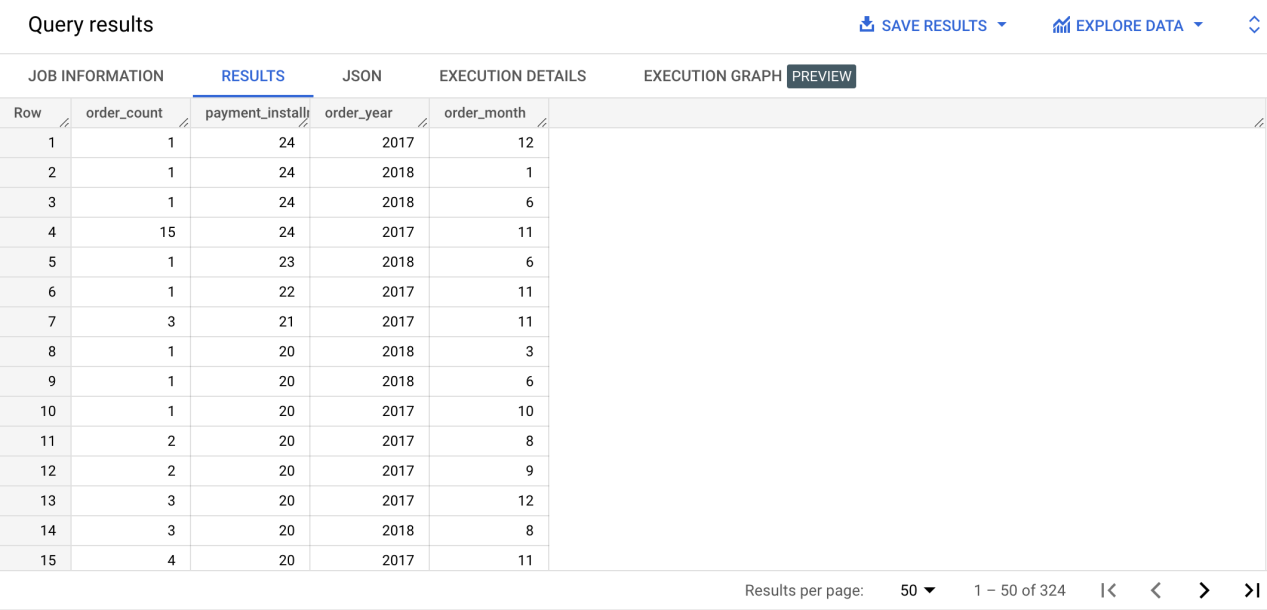
,order\_month

FROM

`BC01\_TargetDataSet.PaymentAnalysis`

ORDER BY payment\_installments DESC, order\_count ASC

;



1. **Actionable Insights**
   1. We can see the highest number of orders were placed in November 2017 which would be during the Black Friday.
   2. More customers tend to shop during Afternoon and Night times.
   3. Customers from SP,RJ,MG,RS,PR states tend to shop a lot.
   4. Delivery taking more time to reach AC,RO,AM,AP,RR states.
   5. Freight value looks a bit high for RO,RR,PB,AC,PI states
   6. Lot of customers use Credit cards and customers who use vouchers are more as compared to who use debit cards.
   7. Orders with number of installments between 1 to 10 are more.
2. **Recommendations**
   1. Need to improve Offers and vouchers during the months around November where people tend to buy more
   2. Also during Afternoon and Night times
   3. Need to provide more offers to customers from SP,RJ,MG,RS,PR states.
   4. Need to improve the delivery time and to check whether the freight charge can be reduced for AC,RO,AM,AP,RR,PB and PI states
   5. More discounts/interest free purchase experience for customers who place orders with 1 to 10 months of installments.