**TARGET CASE STUDY**

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

--A. Data type of all columns in the “customers” table.

**QUERY-**

SELECT

  \*,

  data\_type

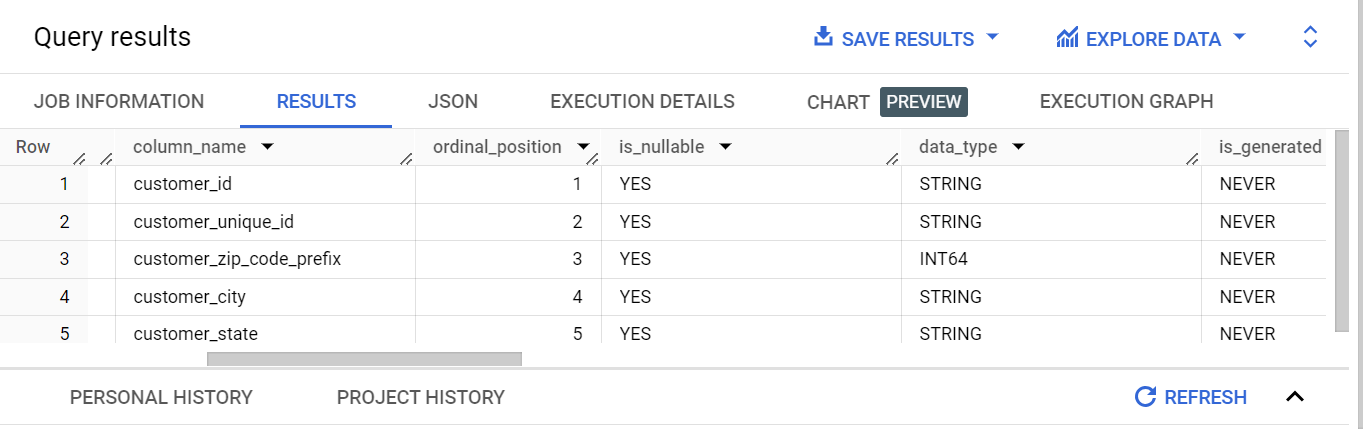
FROM

  `Target.INFORMATION\_SCHEMA.COLUMNS`

WHERE

  table\_name = 'customers';

Screenshot-



Inference-

The customers table has 5 columns with different types as seen above. Columns customer\_id, customer\_unique\_id, customer\_city and customer\_state have data type ‘string’ whereas column customer\_zip\_code\_prefix has data type ‘int64’.

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

**QUERY-**

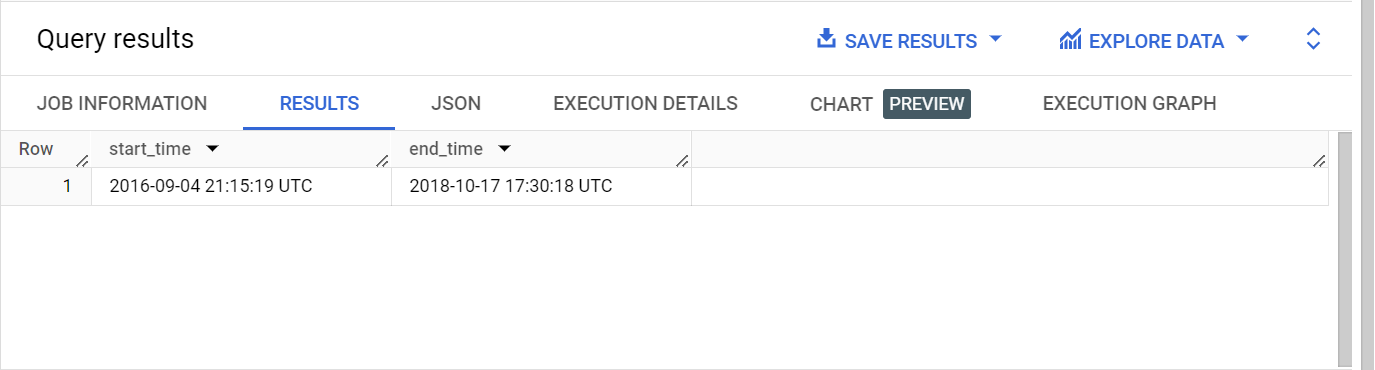
SELECT

MIN(order\_purchase\_timestamp) AS start\_time,

MAX(order\_purchase\_timestamp) AS end\_time

FROM`Target.orders`;

Screenshot-



Inference-

In above image start\_time indicates the timestamp when first order was placed which is in September 2016 where as end\_time indicates the timestamp when the last order was placed which is in October 2018.

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

**QUERY-**

WITH OrdersDuringPeriod AS

(SELECT

customer\_id,

order\_purchase\_timestamp

from `Target.orders`

where order\_purchase\_timestamp between TIMESTAMP('2016-09-04') AND TIMESTAMP('2018-10-17')

),

CustomersInPeriod AS(

  SELECT

  o.customer\_id,

  c.customer\_city,

  c.customer\_state

  FROM

  OrdersDuringPeriod AS o

  JOIN `Target.customers` AS c

  ON o.customer\_id = c.customer\_id

)

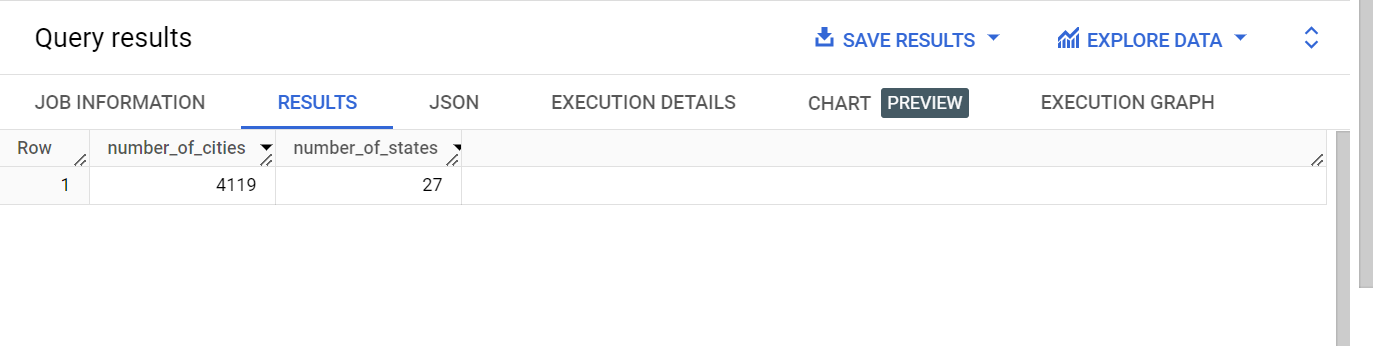
SELECT

COUNT(Distinct(customer\_city)) AS number\_of\_cities,

COUNT(Distinct(customer\_state)) AS number\_of\_states

FROM CustomersInPeriod;

Screenshot-



Inference-

Customers ordered from total 4119 cities in 27 states during given time period.

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

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

**QUERY-**

SELECT

DATE\_TRUNC(order\_purchase\_timestamp, year) AS year,

COUNT(\*) AS order\_count

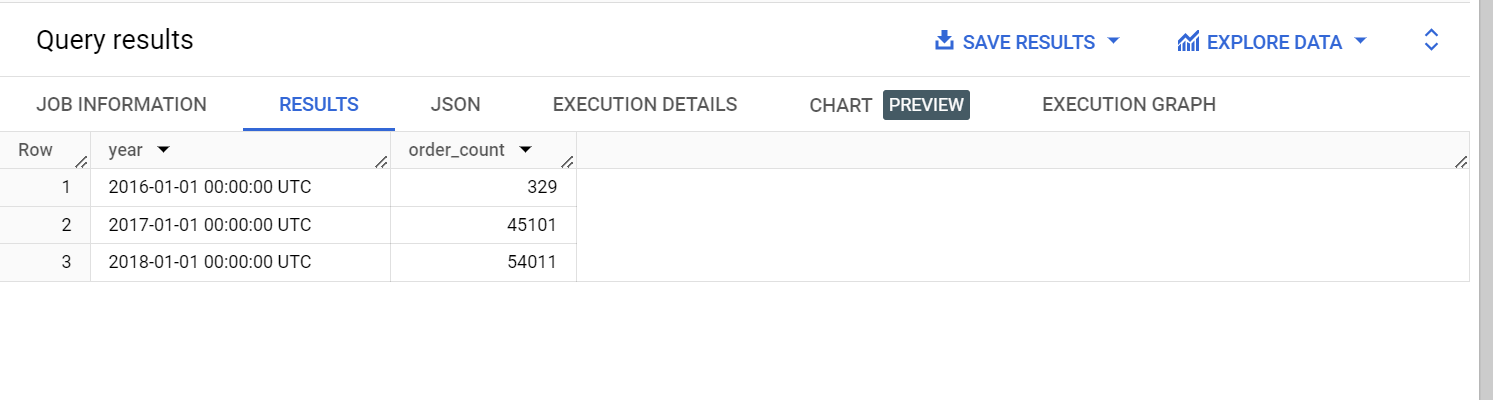
FROM

`Target.orders`

GROUP BY year

ORDER BY year;

Screenshot-



Inference-

From the results it can be observed that number of orders placed shows growing trend over the period of 2016 to 2018 steadily throughout.

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

**QUERY-**

SELECT

DATE\_TRUNC(order\_purchase\_timestamp, month) AS month,

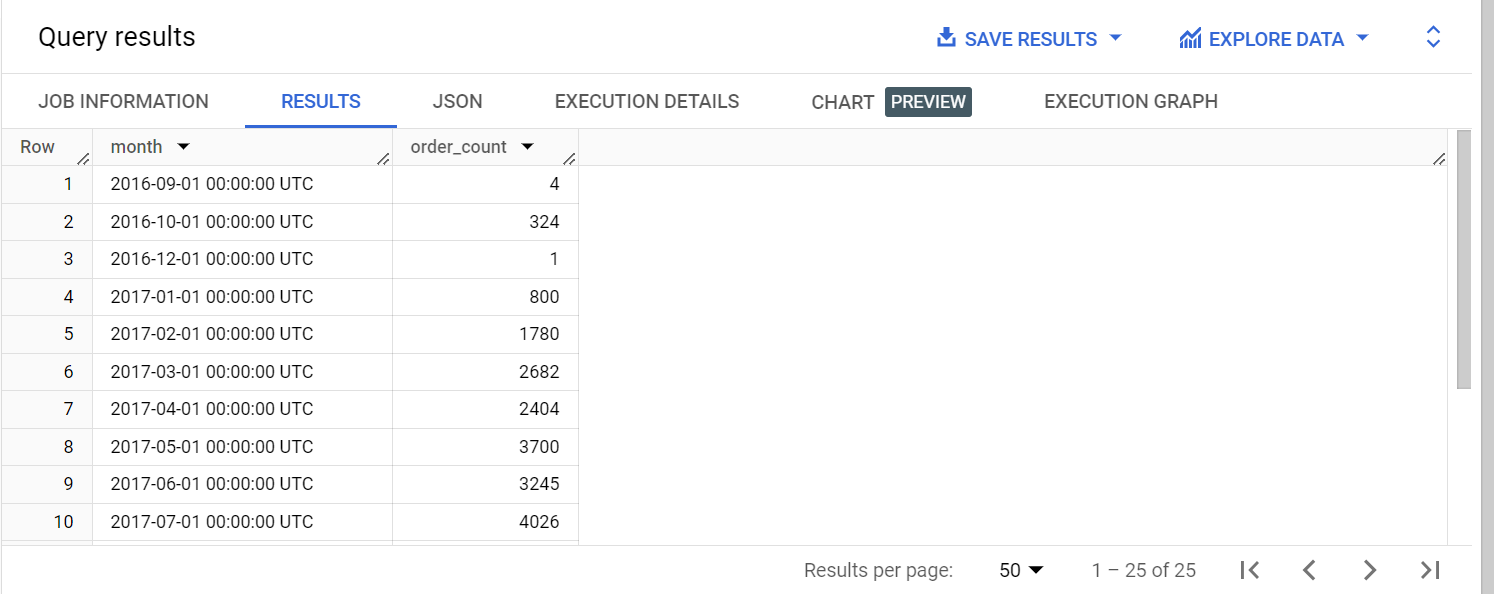
COUNT(\*) AS order\_count

FROM `Target.orders`

GROUP BY month

ORDER BY month;

Screenshot-



Inference-

From results highest number of orders placed were in November 2017 were as lowest numbers of orders placed were in November 2016. During the months of September and October in all three consecutive years 2016, 2017 and 2018, number of orders placed showed decreasing trend.

--C. During what time of the day, do the Brazilian customers mostly place their orders? (Dawn, Morning, Afternoon or Night)

**QUERY-**

WITH BrazilianOrders AS

(SELECT

order\_purchase\_timestamp

FROM

`Target.orders`

)

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 'Mornings'

WHEN EXTRACT(HOUR FROM order\_purchase\_timestamp) >=13 AND EXTRACT(HOUR FROM order\_purchase\_timestamp) <18 THEN 'Afternoon'

ELSE 'Night'

END AS time\_period,

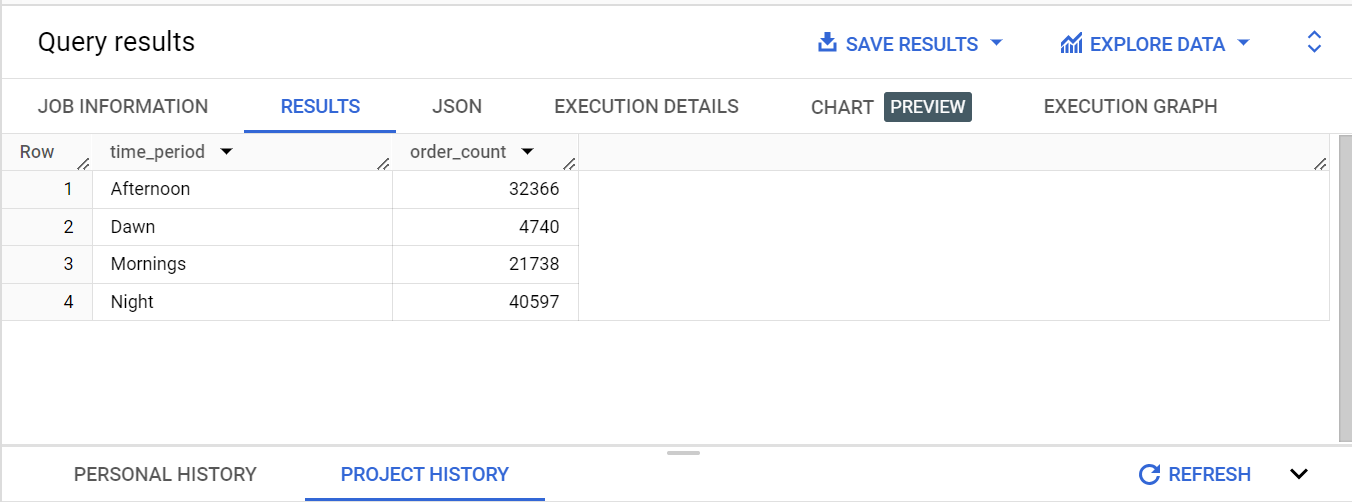
COUNT(\*) as order\_count

FROM BrazilianOrders

GROUP BY time\_period

ORDER BY time\_period;

Screenshot-



Inference-

From the above image it can be seen that Brazilian customers mostly preferred the Night time to place their orders which was between 6 pm and 12 am.

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

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

**QUERY-**

WITH orderCounts AS

(SELECT

EXTRACT(month FROM order\_purchase\_timestamp) AS month,

EXTRACT(year FROM order\_purchase\_timestamp) AS year,

COUNT(\*) AS order\_count

FROM `Target.orders`

GROUP BY month,year

)

SELECT

oc.month,

oc.year,

cs.customer\_state,

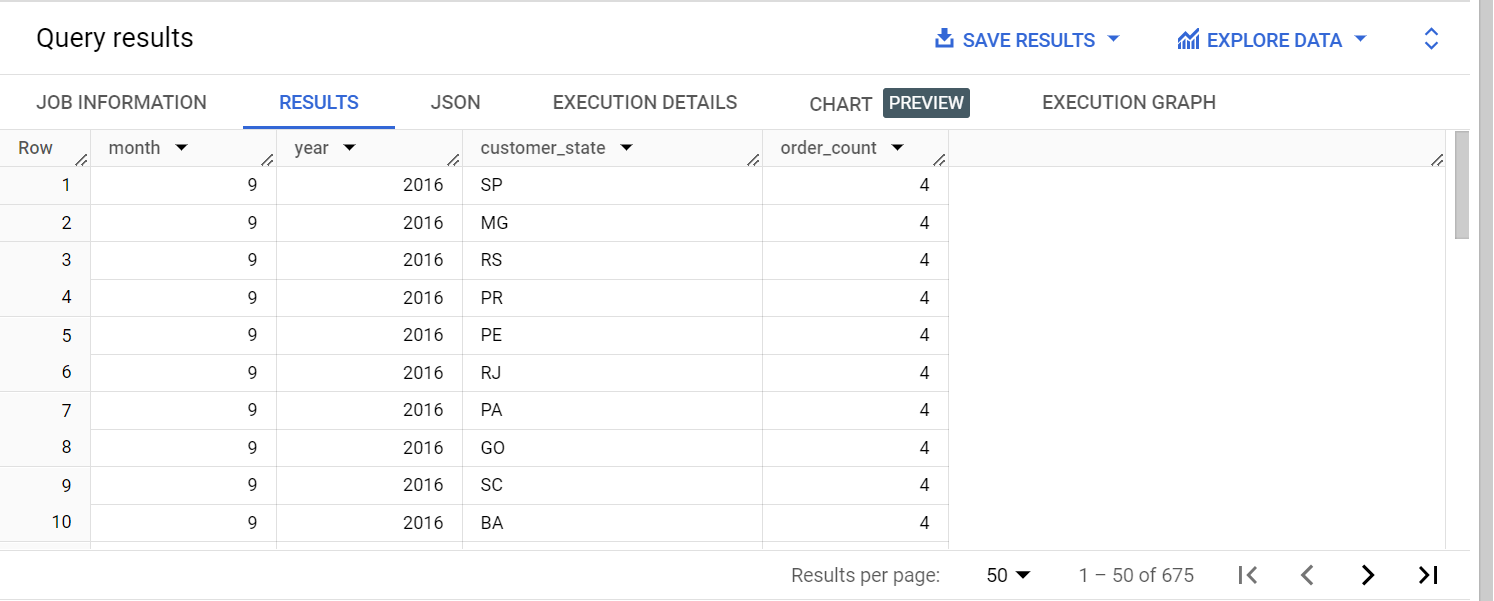
oc.order\_count

FROM orderCounts AS oc

CROSS JOIN `Target.customers` AS cs

GROUP BY month,year,cs.customer\_state,oc.order\_count;

Screenshot-



Inference-

Month on month orders placed in each state can be seen in above image.

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

**QUERY-**

SELECT

customer\_state,

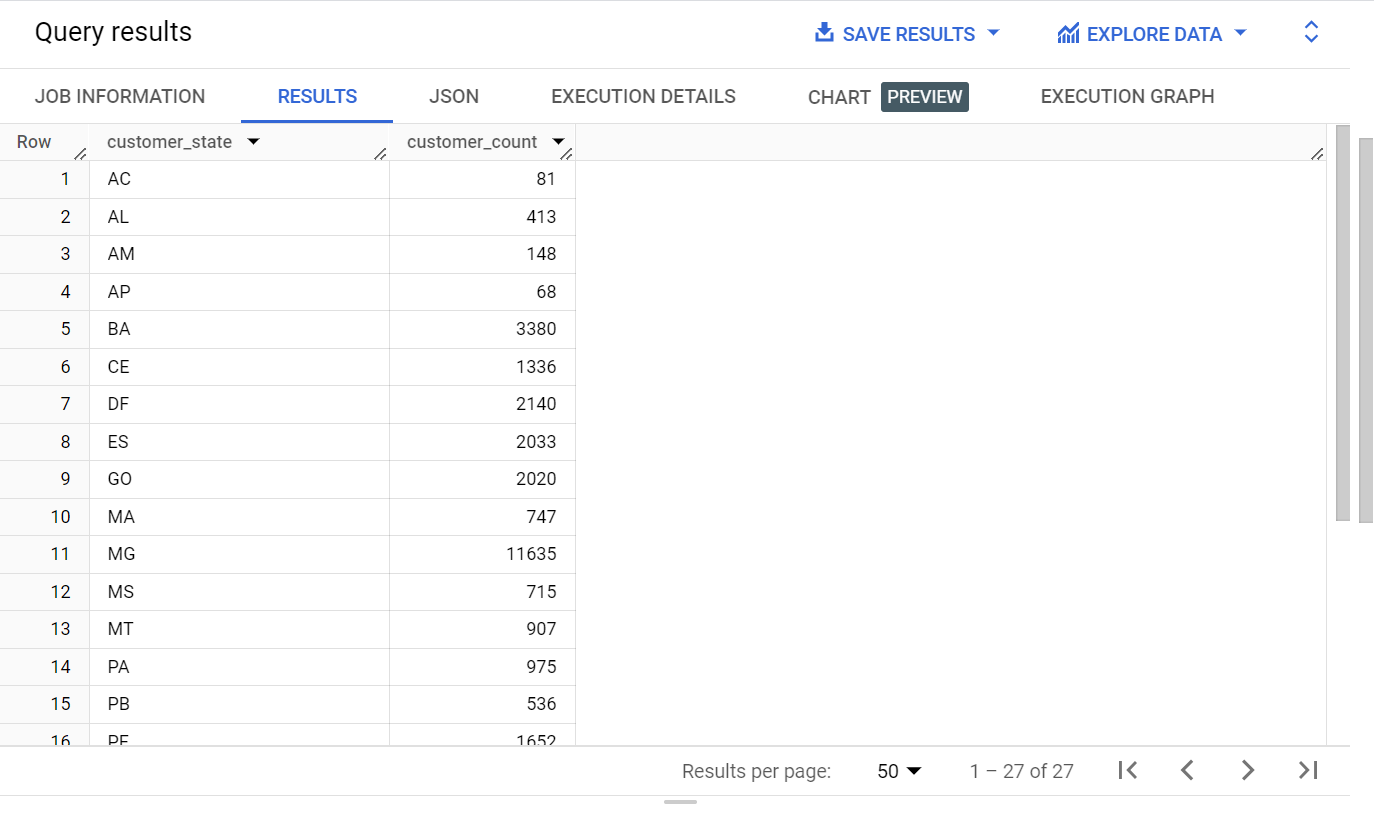
COUNT(DISTINCT customer\_id) as customer\_count

FROM `Target.customers`

GROUP BY customer\_state

ORDER BY customer\_state;

Screenshot-



Inference-

AS seen above from total 27 states, states of São Paulo (SP), Rio de Janeiro (RJ) and Minas Gerais (MG) have highest concentration of customers respectively. Average customer count is 907.

**--4. Impact on Economy: Analyze the money movement by e-commerce by looking at order prices, freight and others.**

--A. Get the % increase in the cost of orders from year 2017 to 2018 (include months between Jan to Aug only).

**QUERY-**

WITH monthlyCost AS

(

  SELECT

  EXTRACT(month FROM o.order\_purchase\_timestamp) as month,

  EXTRACT(year FROM o.order\_purchase\_timestamp) as year,

  SUM(oi.price) as total\_cost

  FROM `Target.orders` AS o

  FULL JOIN `Target.order\_items` AS oi

  on oi.order\_id = o.order\_id

  WHERE EXTRACT(month FROM o.order\_purchase\_timestamp) BETWEEN 1 and 8

  AND EXTRACT(year FROM o.order\_purchase\_timestamp) in (2017,2018)

  GROUP BY month,year

  ORDER BY year,month

)

SELECT

m1.year,

m1.month,

(m2.total\_cost-m1.total\_cost)/m1.total\_cost \* 100 AS percent\_increase

FROM monthlyCost AS m1

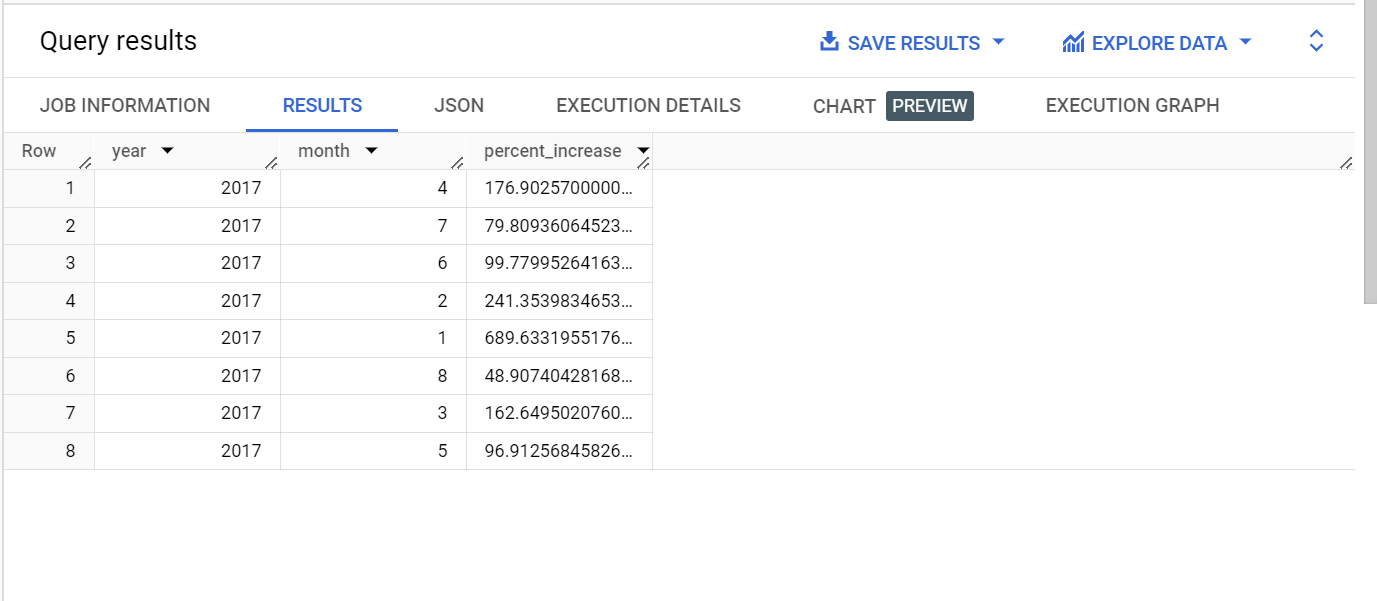
JOIN monthlyCost AS m2

ON m1.year = 2017

AND m2.year = 2018

AND m1.month = m2.month;

Screenshot-



Inference-

From the image above it can be seen that most increment in order prices had occurred in January 2018 as compared to January 2017 by 689.64 percent. Followed by February 2018 and April 2018 with 242 percent and 177 percent respectively. Lowest increment can be seen in August 2018 with 48.9 percent increment.

--B. Calculate the Total & Average value of order price for each state.

**QUERY-**

SELECT

c.customer\_state,

SUM(oi.price) AS total\_cost,

AVG(oi.price) AS avg\_cost

FROM `Target.customers` AS c

JOIN `Target.orders` AS o

ON c.customer\_id = o.customer\_id

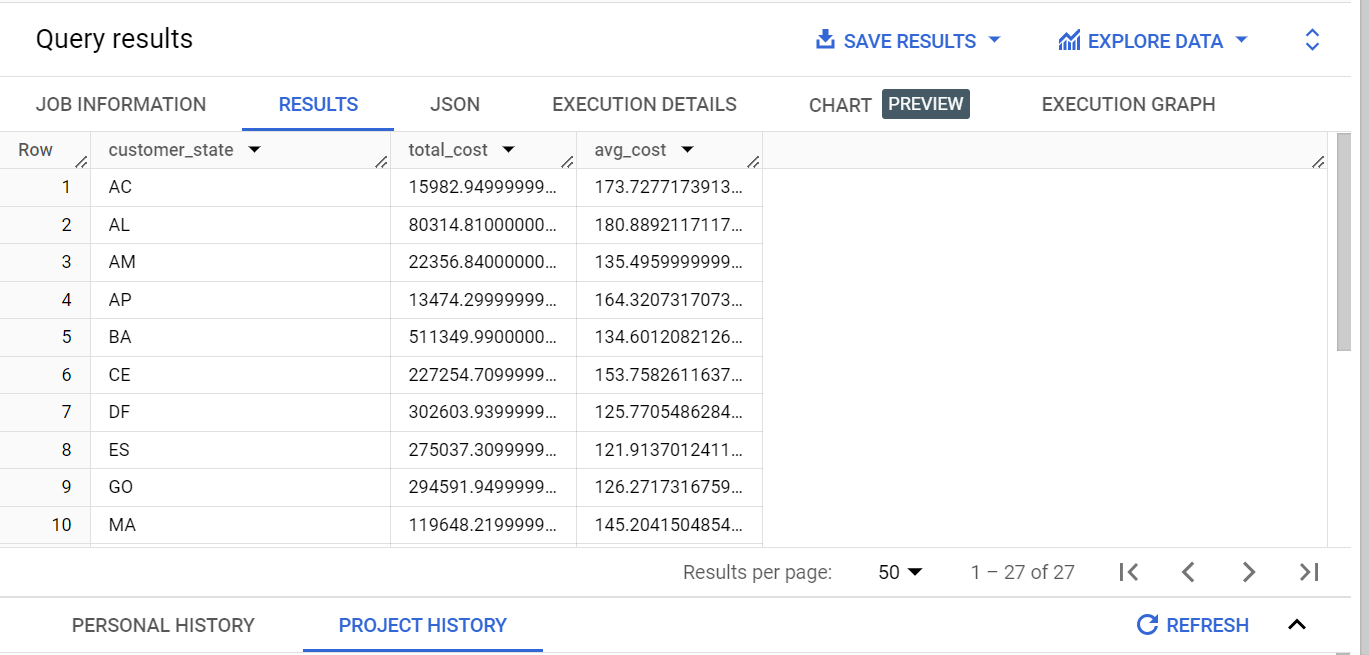
JOIN `Target.order\_items` AS oi

ON o.order\_id = oi.order\_id

GROUP BY c.customer\_state

ORDER BY c.customer\_state;

Screenshot-



Inference-

Highest average order price is in the state of Paraíba and lowest average order price is in state of São Paulo.

--C. Calculate the Total & Average value of order freight for each state.

**QUERY-**

SELECT

c.customer\_state,

SUM(oi.freight\_value) AS total\_freight\_value,

AVG(oi.freight\_value) AS avg\_freight\_value

FROM `Target.orders` AS o

JOIN `Target.order\_items` AS oi

ON o.order\_id = oi.order\_id

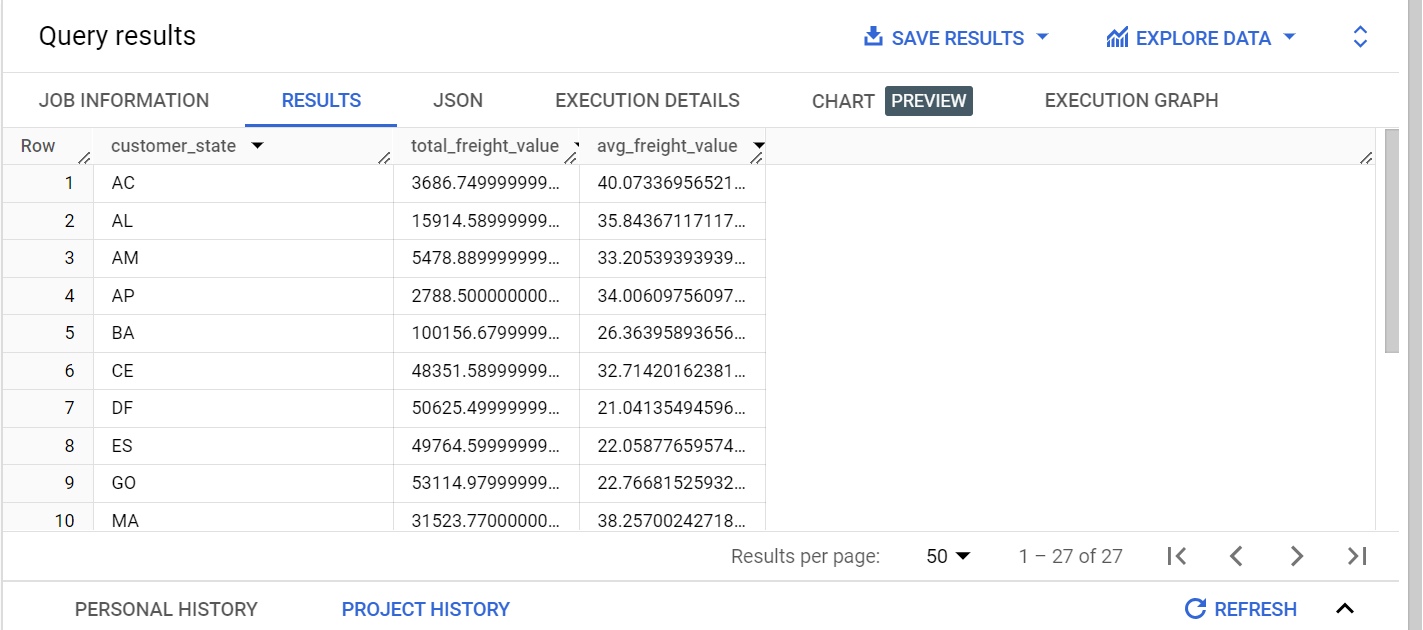
JOIN `Target.customers` AS c

ON c.customer\_id = o.customer\_id

GROUP BY c.customer\_state

ORDER BY c.customer\_state;

Screenshot-



Inference-

Total & Average value of order freight for each state is shown in above image. Highest average order freight value is in the state of Roraima and lowest average order freight value is in state of São Paulo. Total value of order freight for these 2 states is 2235.19 and 718723.06 respectively.

**--5. Analysis based on sales, freight and delivery time**

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

Do this in a single query.

**QUERY-**

SELECT

order\_id,

order\_purchase\_timestamp,

order\_delivered\_customer\_date,

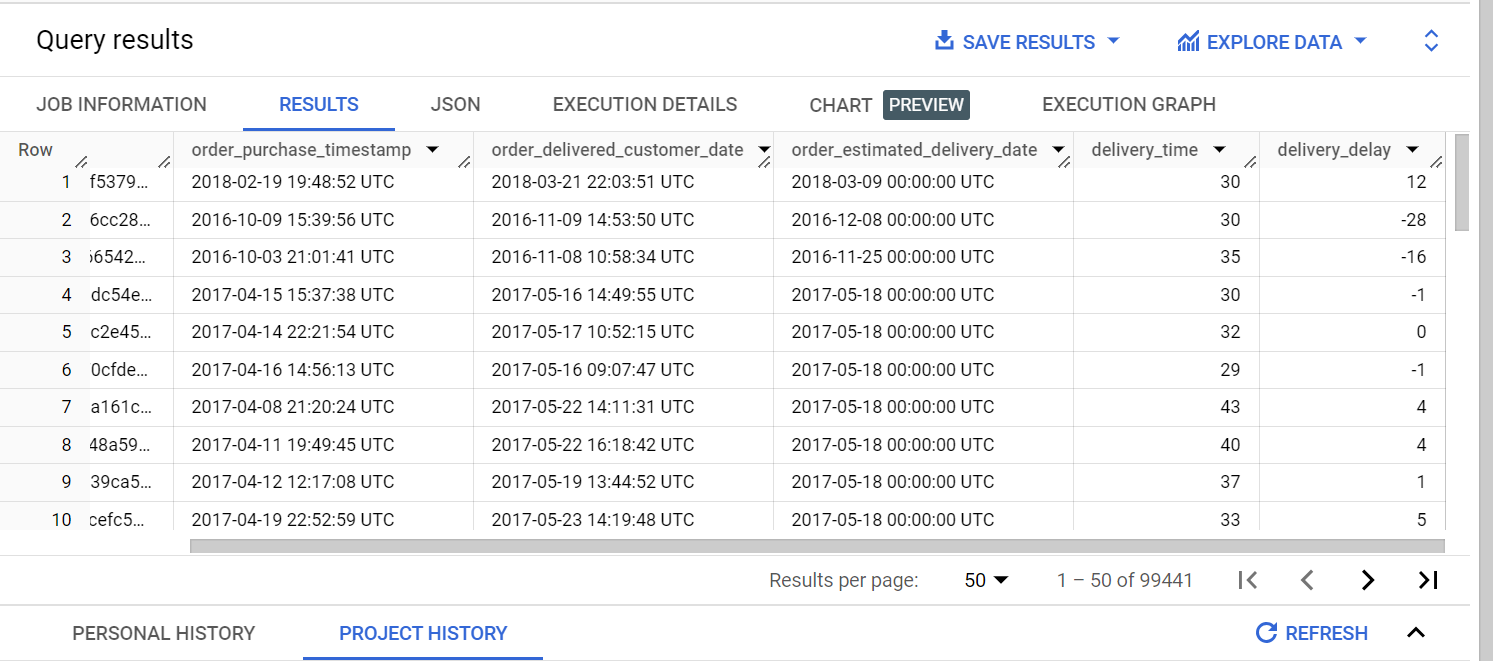
order\_estimated\_delivery\_date,

DATE\_DIFF(order\_delivered\_customer\_date, order\_purchase\_timestamp, day) AS delivery\_time,

DATE\_DIFF(order\_delivered\_customer\_date, order\_estimated\_delivery\_date,day) AS delivery\_delay

FROM `Target.orders`;

Screenshot-



Inference-

Number of days taken to deliver each order from the order’s purchase date and difference (in days) between the estimated & actual delivery date of an order is calculated under delivery time and delivery delay column for each state. Longest delivery time was of 67 days and shortest was of 29 days. Longest delivery delay was of 67 days and shortest delay was of 28 days before the estimated delivery date.

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

**QUERY-**

SELECT

c.customer\_state,

AVG(oi.freight\_value) AS avg\_freight\_value

FROM `Target.orders` as o

JOIN `Target.order\_items` AS oi

ON o.order\_id = oi.order\_id

JOIN `Target.customers` AS c

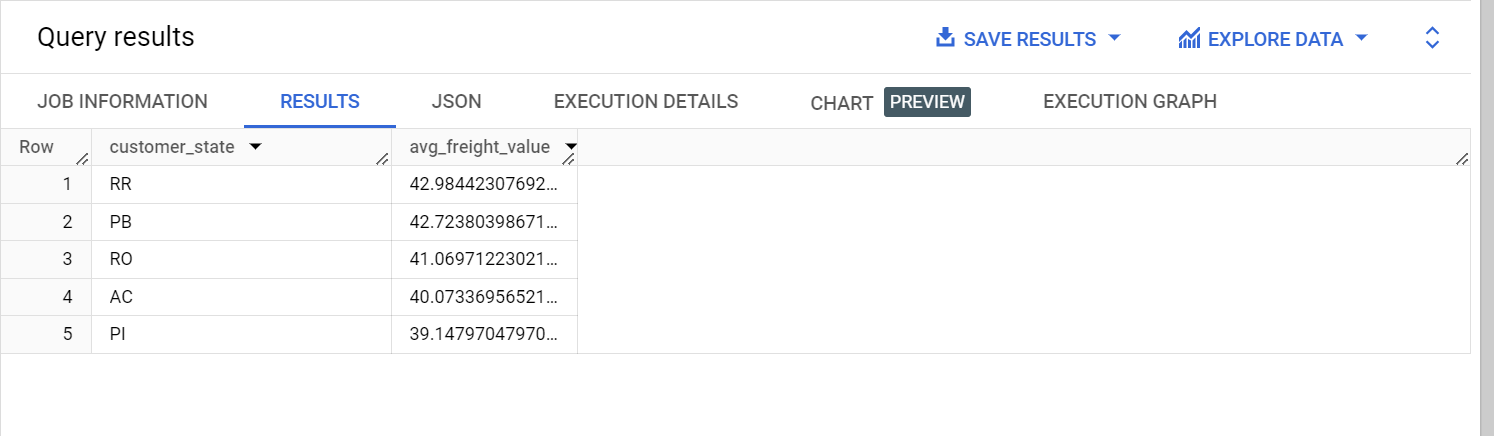
ON c.customer\_id = o.customer\_id

GROUP BY c.customer\_state

ORDER BY avg\_freight\_value DESC

LIMIT 5;

Screenshot-



Inference-

Top 5 states with highest average freight values are Roraima, Paraíba, Rondônia, Acre and Piauí.

SELECT

c.customer\_state,

AVG(oi.freight\_value) AS avg\_freight\_value

FROM `Target.orders` as o

JOIN `Target.order\_items` AS oi

ON o.order\_id = oi.order\_id

JOIN `Target.customers` AS c

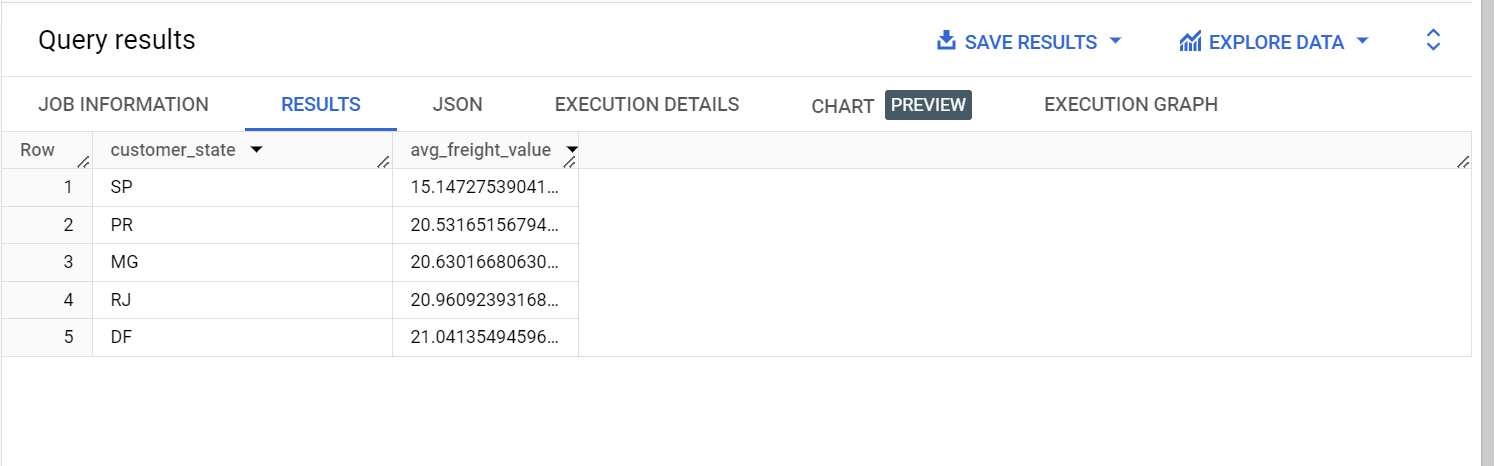
ON c.customer\_id = o.customer\_id

GROUP BY c.customer\_state

ORDER BY avg\_freight\_value

LIMIT 5;

Screenshot-



Inference-

Top 5 states with lowest average freight values are São Paulo, Paraná, Minas Gerais, Rio de Janeiro and Distrito Federal.

--c. Find out the top 5 states with the highest & lowest average delivery time.

**QUERY-**

SELECT

c.customer\_state,

AVG(DATE\_DIFF(o.order\_delivered\_customer\_date, o.order\_purchase\_timestamp, day)) AS avg\_deliveryTime

FROM `Target.customers` AS c

JOIN `Target.orders` AS o

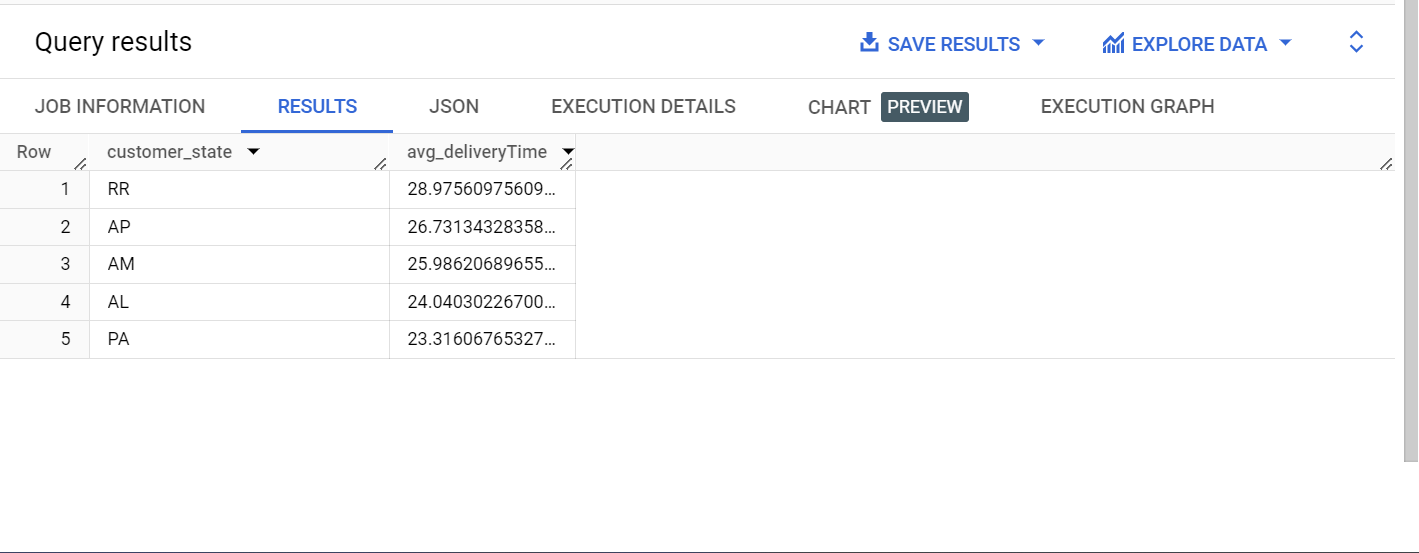
ON c.customer\_id = o.customer\_id

GROUP BY c.customer\_state

ORDER BY avg\_deliveryTime DESC

LIMIT 5;

Screenshot-



Inference-

Top 5 states with the highest average delivery time are Roraima, Amapá, Amazonas, Alagoas and Pará.

SELECT

c.customer\_state,

AVG(DATE\_DIFF(o.order\_delivered\_customer\_date, o.order\_purchase\_timestamp, day)) AS avg\_deliveryTime

FROM `Target.customers` AS c

JOIN `Target.orders` AS o

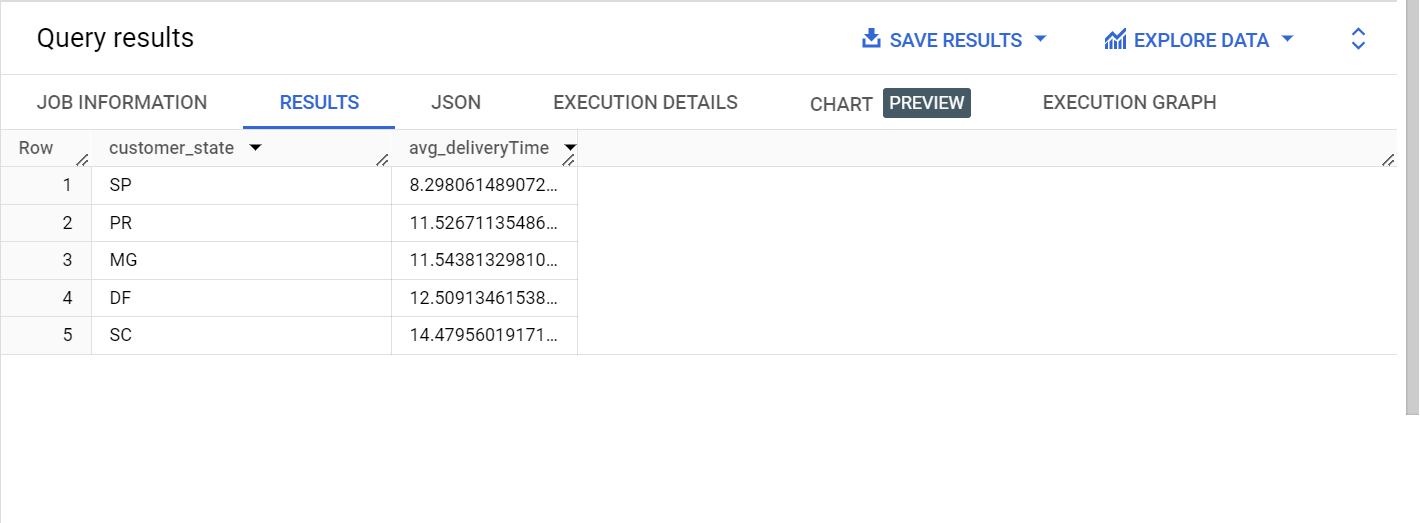
ON c.customer\_id = o.customer\_id

GROUP BY c.customer\_state

ORDER BY avg\_deliveryTime

LIMIT 5;

Screenshot-



Inference-

Top 5 states with the lowest average delivery time are São Paulo, Paraná, Minas Gerais, Distrito Federal and Santa Catarina.

--D. Find out the top 5 states where the order delivery is really fast as compared to the estimated date of delivery.

**QUERY-**

SELECT

c.customer\_state,

AVG(DATE\_DIFF(o.order\_delivered\_customer\_date,o.order\_estimated\_delivery\_date,day)) AS deliveryTime

FROM `Target.customers` AS c

JOIN `Target.orders` AS o

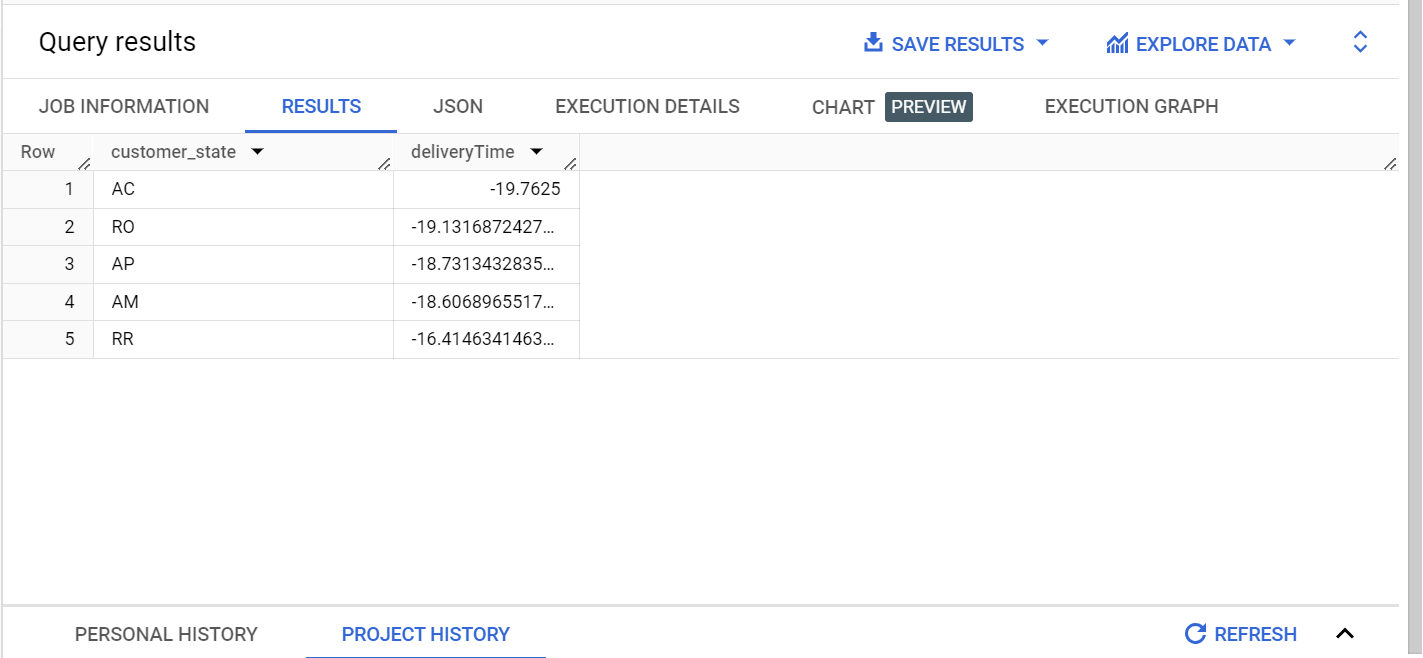
ON c.customer\_id=o.customer\_id

GROUP BY c.customer\_state

ORDER BY deliveryTime

LIMIT 5;

Screenshot-



Inference-

Top 5 states with the fastest order delivery as compared to the estimated date of delivery are Acre, Rondônia, Amapá, Amazonas and Roraima respectively.

**-- 6. Analysis based on the payments:**

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

**QUERY-**

SELECT

EXTRACT(year FROM o.order\_purchase\_timestamp) AS year,

EXTRACT(month FROM o.order\_purchase\_timestamp) AS month,

p.payment\_type,

COUNT(\*) AS order\_count

FROM `Target.payments` AS p

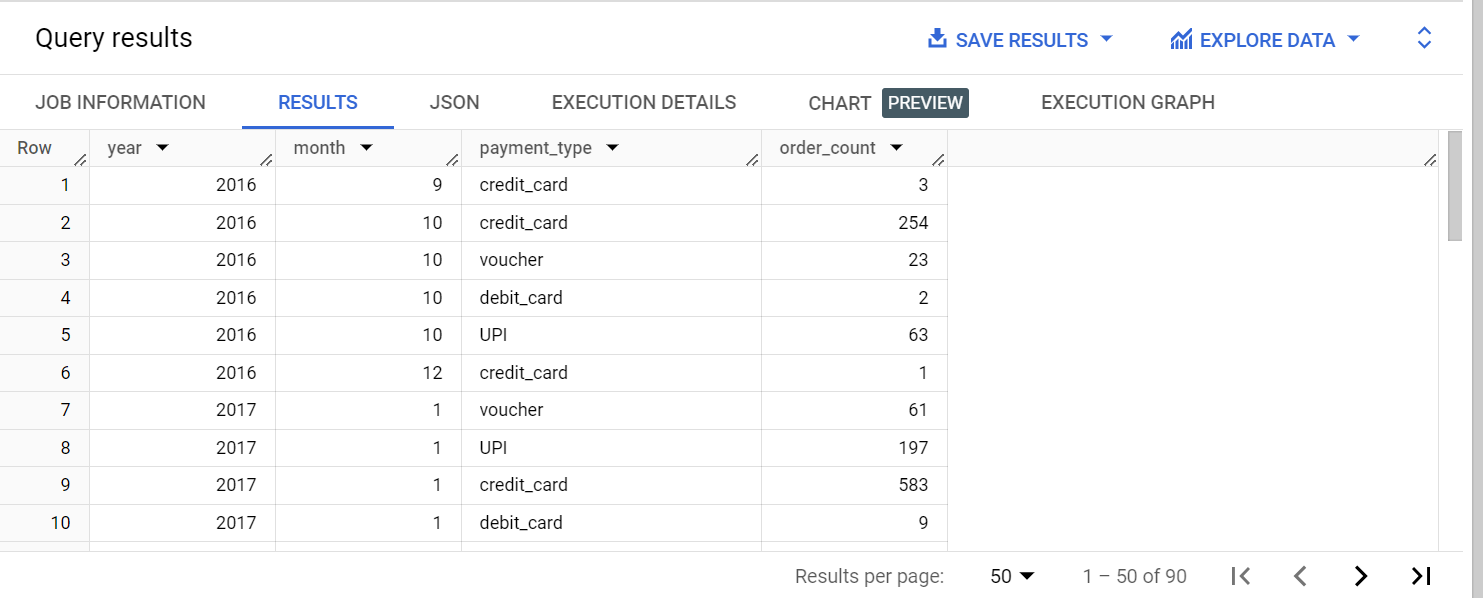
JOIN `Target.orders` AS o

ON p.order\_id = o.order\_id

GROUP BY month, year, p.payment\_type

ORDER BY month,year;

Screenshot-



Inference-

In above image we can see count of month on month of all payment types under the column order count. From the results it can be seen most of the orders in each month are placed via credit card payment type followed by UPI payment system.

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

**QUERY-**

SELECT

payment\_installments,

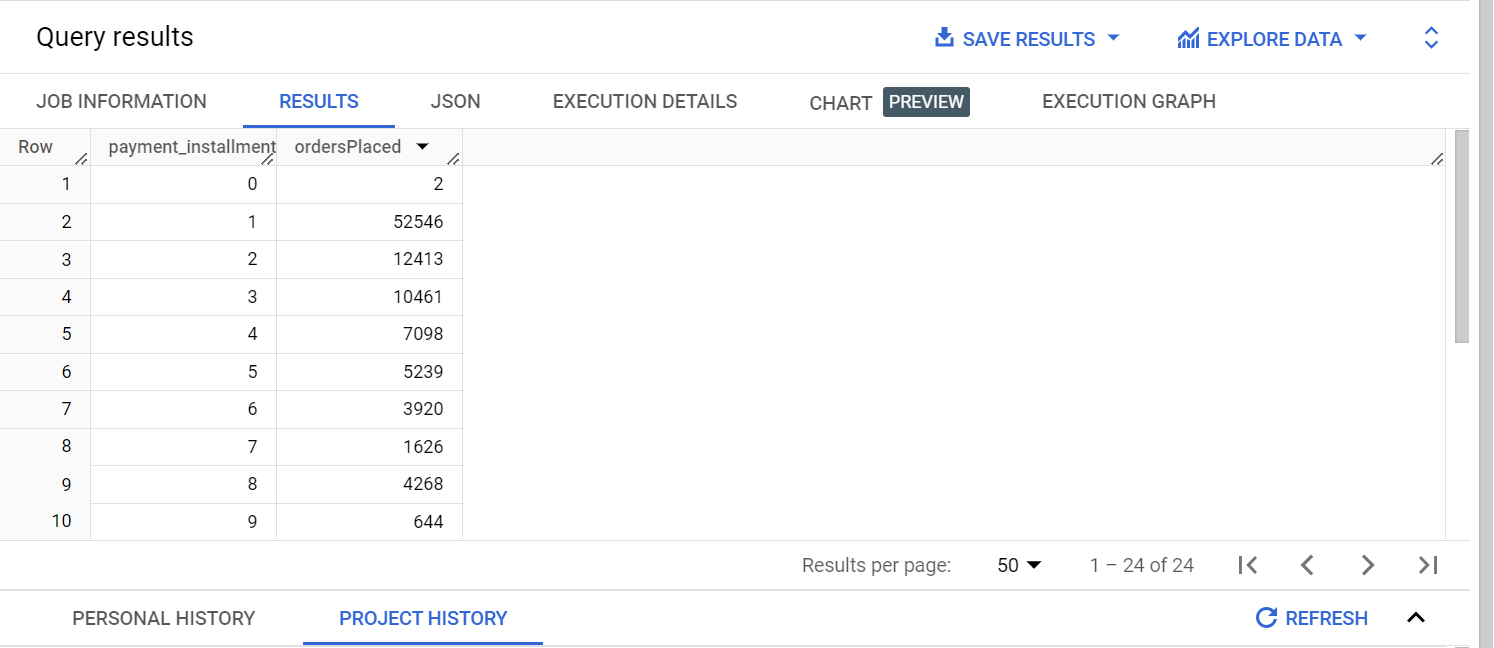
COUNT(\*) AS ordersPlaced

FROM `Target.payments`

GROUP BY payment\_installments

ORDER BY payment\_installments;

Screenshot-



Inference-

From above image we can see most orders were paid in single instalment, followed by 2 instalments and 3 instalments.

**Insights** –

* SP state (Sao Paulo) has Lowest average freight price and lowest average order price as well.
* It is the same state having lowest average delivery time as well with being one of the most densely populated areas along with Rio de Janeiro (RJ) and Minas Gerais (MG).
* Densely populated areas have lowest average freight values.
* These states have lowest average delivery time and slowest order delivery time as compared to estimated delivery date.
* Interestingly states having Highest average freight price are also the states which have highest average order price like Roraima, Amazonas etc.
* These are same states which have highest average delivery time and fastest order delivery time as compared to estimated delivery date.
* These are the states which have Lower population concentration.
* So we can say densely populated areas are difficult areas resulting in slower order delivery time and higher freight values and hence resulting in lowest average order price.

**Recommendations** –

A reputation for reliable and fast deliveries can positively impact brand perception, influencing purchasing decisions and attracting more customers.

* Improving the average delivery time of orders in poorly performing states like Sao Paulo and Rio de Janeiro, especially for ecommerce giants like Target, involves strategic initiatives and investments. Here are some potential actions that we can take –
* **Optimise Distribution centres**-

Establish strategically located distribution centres within or near metropolitan areas to reduce the distance travelled for deliveries.

* **Implement Regional Warehousing** –

Utilize regional warehousing to store popular or frequently ordered items closer to customers, enabling quicker order fulfilment.

* **Utilize Advanced Technology** –

Implement advanced technologies such as warehouse automation, robotics, and artificial intelligence to streamline order processing and reduce handling times.

* **Dynamic Routing and Tracking –**

Implement dynamic routing algorithms and real-time tracking systems to optimize delivery routes and provide customers with accurate delivery estimates.

* **Collaborate with Local Logistics Providers –**

Establish partnerships with local logistics providers who have a deep understanding of the local infrastructure, traffic patterns, and regulatory requirements.

* **Incentivize Off-Peak Deliveries –**

Encourage customers to choose off-peak delivery times by offering discounts or incentives, which can help reduce congestion and improve delivery efficiency.