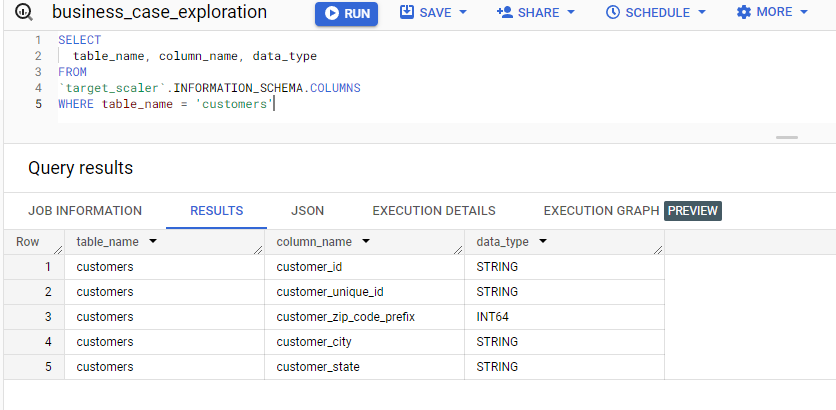
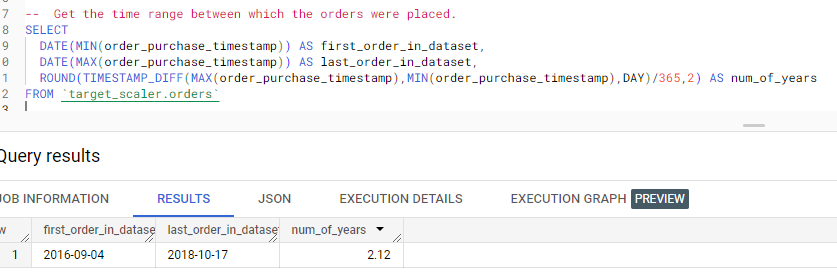
TARGET BUSINESS CASE PROJECT

Submitted by – Shahzar Husain

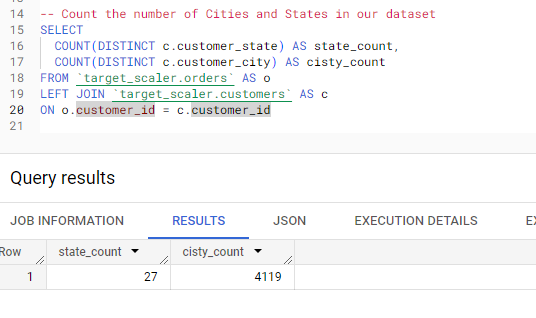
1. **Importing the dataset and do usual exploratory analysis steps like checking the structure & characteristics of the dataset:**
   1. **Data type of all columns in the "customers" table.**



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



* 1. **Check the number of Cities and States in our dataset.**



**INSIGHTS -**

In the basic exploration, we checked and the data seems to have stored with the correct data type. Additionally, the data is only for a period of 26 months that is spread across 3 years and as the third query shows, our data has customers from 27 states and approx. 4k cities.

1. **In-depth Exploration to check:**  
   1. **Is there a growing trend in the no. of orders placed over the past years?**

****

**INSIGHTS –** There is a visible growth in the number of orders placed each year. We have data since Sept’16 that’s the reason of such low number of orders in the year 2016.

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

**QUERY:**

SELECT

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

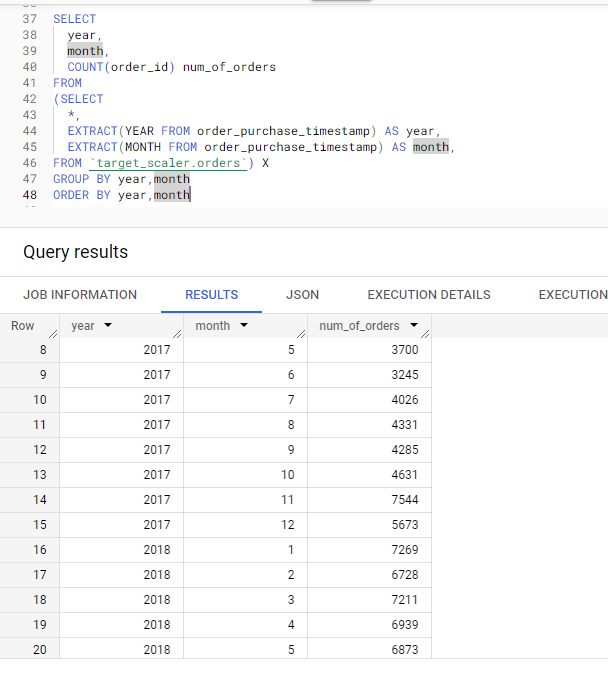
  EXTRACT(MONTH FROM order\_purchase\_timestamp) AS month,

  COUNT(\*)

FROM `target\_scaler.orders`

GROUP BY year,month

ORDER BY year,month

****

**INSIGHTS –** We can see a growing trend for number of orders month-on-month. However, the data does not seem sufficient enough to comment on the seasonality, as we have only 4th quarter data for 2016 but no 4th quarter data is present for 2018.

The only dip in number of orders we can see is in the month of December for both 2016 and 2017.

* 1. **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

**QUERY:**

SELECT

  time\_of\_day,

  COUNT(order\_id) AS num\_of\_orders

FROM

(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 'Morning'

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

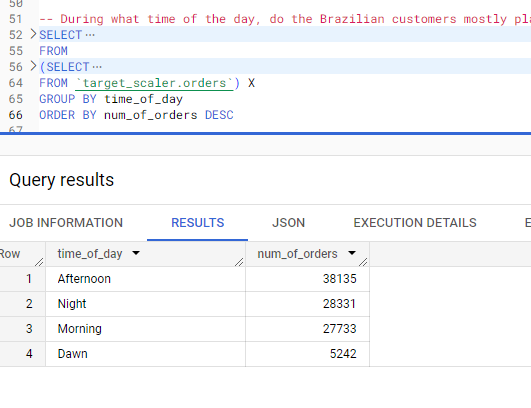
    ELSE 'Night'

    END as time\_of\_day

FROM `target\_scaler.orders`) X

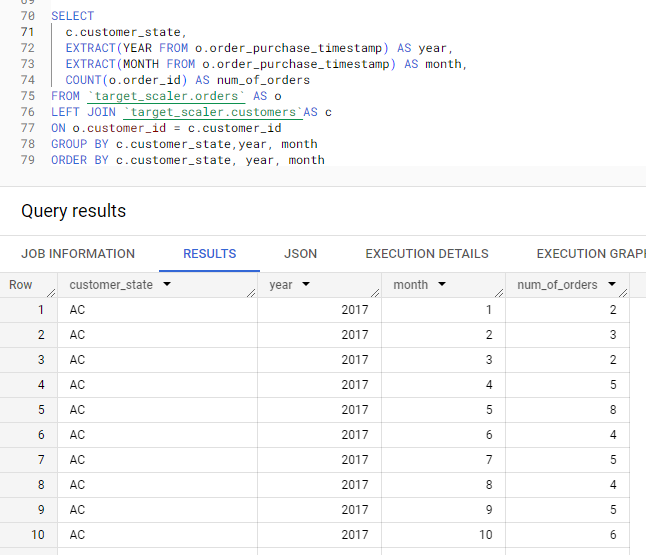
GROUP BY time\_of\_day

ORDER BY num\_of\_orders DESC

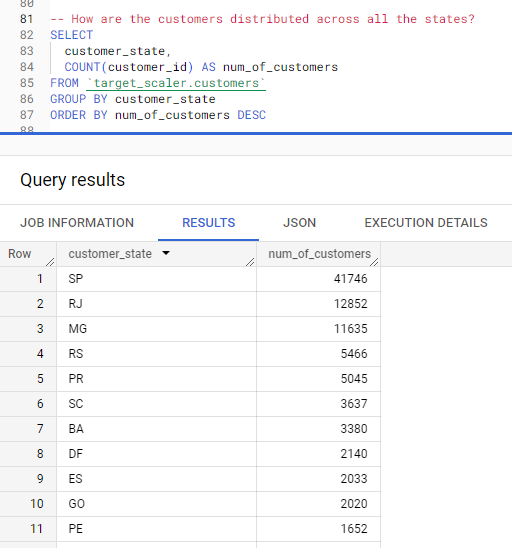
****

**INSIGHTS –** It seems that majority of Brazilians do like to party and sleep during the night and they are up for shopping in the afternoon time.

1. **Evolution of E-commerce orders in the Brazil region:**
   1. **Get the month-on-month no. of orders placed in each state.**

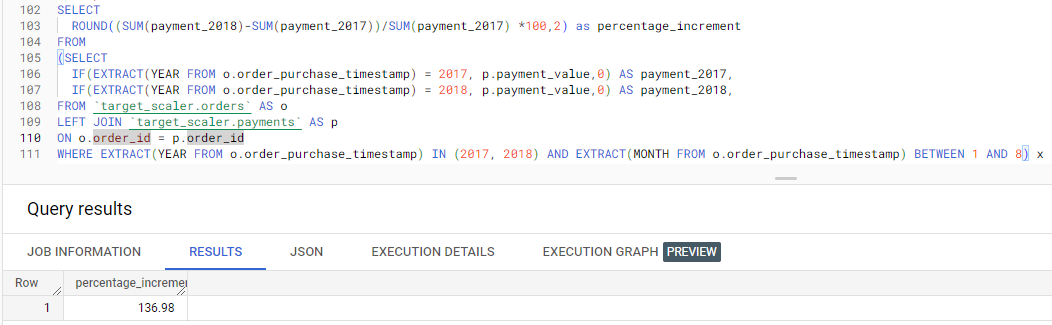
****

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

****

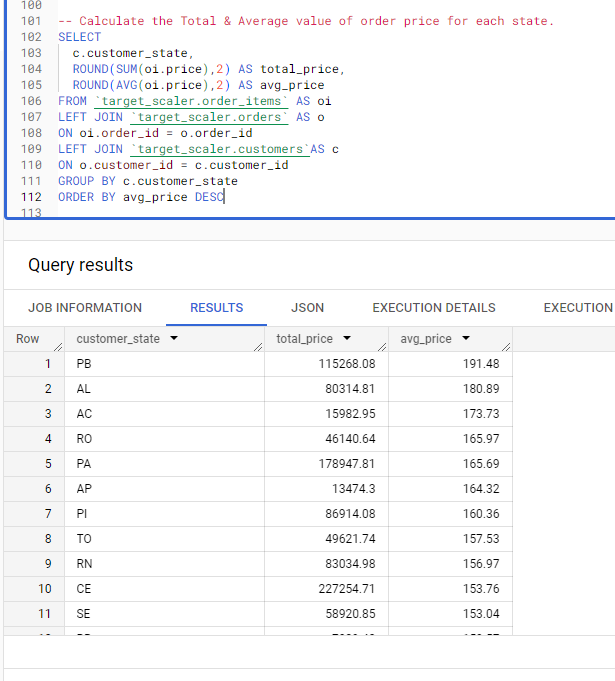
**INSIGHTS –** Approx40% of our customer base is from Brazil’s capital, SP i.e., São Paulo.

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

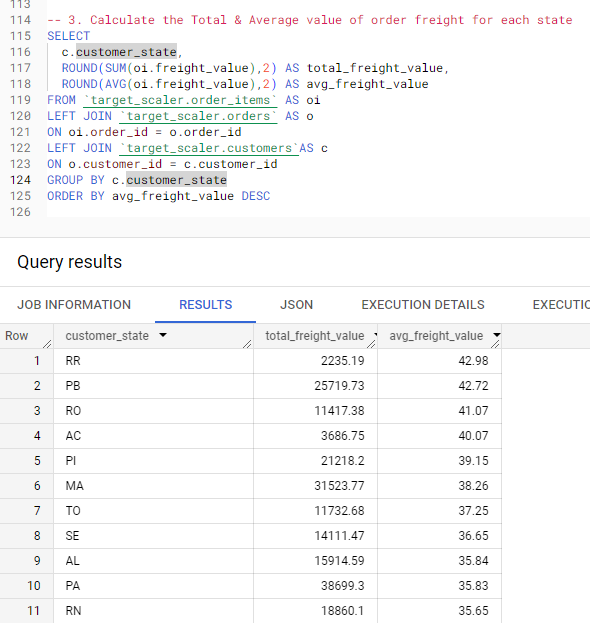


**INSIGHTS –** We can see a whopping ~137% increase in purchase value from 2017 to 2018.

* 1. **Calculate the Total & Average value of order price for each state.**

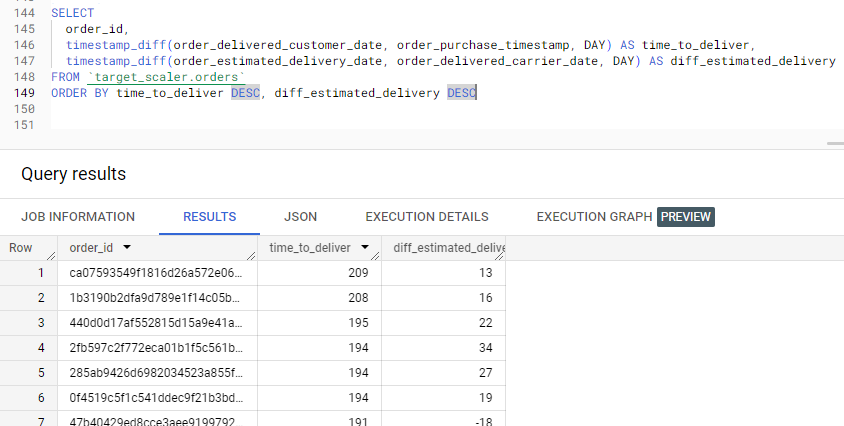
****

* 1. **Calculate the Total & Average value of order freight for each state.**

****

**INSIGHTS -** We can see that State of Paraiba has the highest average order price among all the states and RR i.e., State of Roraima has the highest average value for freight.

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.

****

**INSIGHTS –** As far as the delivery is concerned, with above information, we can see that it takes around 6 months as well for some order to deliver and the difference between the estimated delivery and actual can differ from few days to a month

* 1. **Find out the top 5 states with the highest & lowest average freight value.**

**QUERY:**

SELECT

  c.customer\_state,

  ROUND(AVG(oi.freight\_value),2) AS top\_avg\_freight\_value

FROM `target\_scaler.order\_items` AS oi

LEFT JOIN `target\_scaler.orders` AS o

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

LEFT JOIN `target\_scaler.customers`AS c

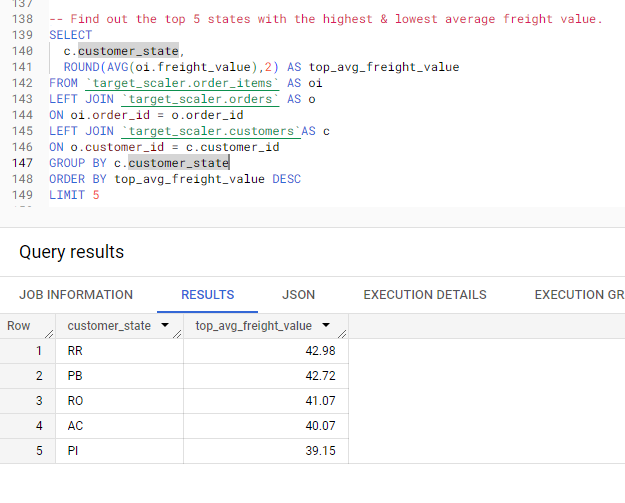
ON o.customer\_id = c.customer\_id

GROUP BY c.customer\_state

ORDER BY top\_avg\_freight\_value DESC

LIMIT 5

**5 states with highest freight value:**

****

**5 states with lowest freight value:**

**QUERY:**

SELECT

  c.customer\_state,

  ROUND(AVG(oi.freight\_value),2) AS top\_avg\_freight\_value

FROM `target\_scaler.order\_items` AS oi

LEFT JOIN `target\_scaler.orders` AS o

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

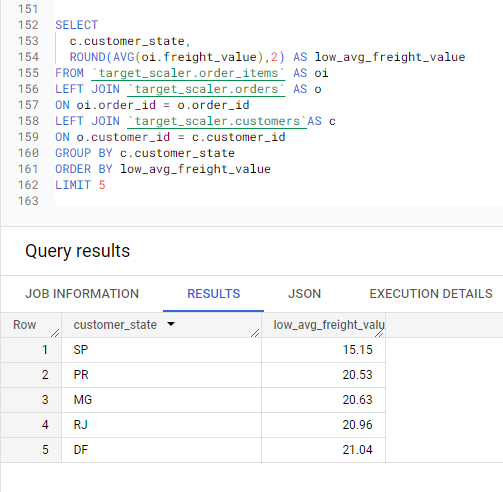
LEFT JOIN `target\_scaler.customers`AS c

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

GROUP BY c.customer\_state

ORDER BY top\_avg\_freight\_value DESC

LIMIT 5

****

**INSIGHTS -** RR has the highest avg\_freight value on the other hand SP has the lowest

* 1. Find out the top 5 states with the highest & lowest average delivery time.

**5 states with lowest delivery time:**

SELECT

  c.customer\_state,

  ROUND(AVG(timestamp\_diff(o.order\_delivered\_customer\_date, o.order\_purchase\_timestamp, DAY)),2) AS delievery\_time

FROM `target\_scaler.order\_items` AS oi

LEFT JOIN `target\_scaler.orders` AS o

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

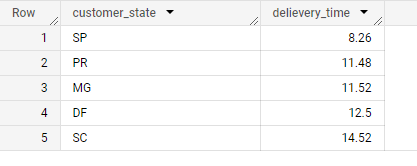
LEFT JOIN `target\_scaler.customers`AS c

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

GROUP BY c.customer\_state

ORDER BY delievery\_time

LIMIT 5



**5 STATES WITH HIGHEST AVEGRAGE DELIEVERY TIME:**

SELECT

  c.customer\_state,

  ROUND(AVG(timestamp\_diff(o.order\_delivered\_customer\_date, o.order\_purchase\_timestamp, DAY)),2) AS delievery\_time

FROM `target\_scaler.order\_items` AS oi

LEFT JOIN `target\_scaler.orders` AS o

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

LEFT JOIN `target\_scaler.customers`AS c

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

GROUP BY c.customer\_state

ORDER BY delievery\_time DESC

LIMIT 5



**INSIGHTS –** SP has the lowest avg delivery time and RR has the highest

* 1. 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,

  ROUND(AVG(timestamp\_diff(o.order\_estimated\_delivery\_date,o.order\_delivered\_customer\_date, DAY)),2) AS delievery\_time

FROM `target\_scaler.order\_items` AS oi

LEFT JOIN `target\_scaler.orders` AS o

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

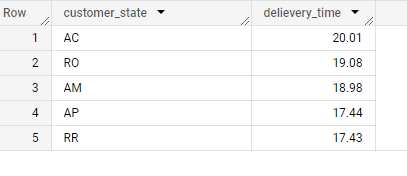
LEFT JOIN `target\_scaler.customers`AS c

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

GROUP BY c.customer\_state

ORDER BY delievery\_time DESC

LIMIT 5

****

**INSIGHTS –** Good news!! If you are living in state of Acre, your order might arrive 20 days early than estimated.

1. **Analysis based on the payments:**
   1. Find the month-on-month no. of orders placed using different payment types.

QUERY:

SELECT

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

  p.payment\_type,

  COUNT(p.order\_id) AS num\_of\_orders

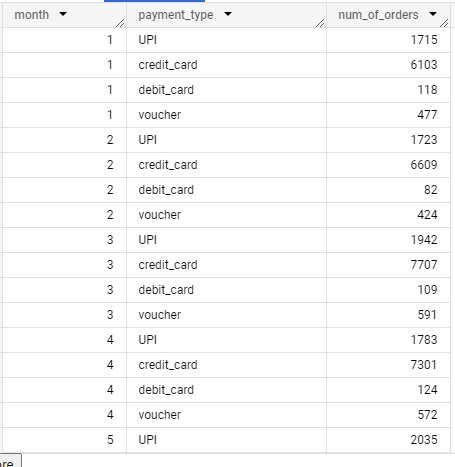
FROM `target\_scaler.payments` AS p

LEFT JOIN `target\_scaler.orders` AS o

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

GROUP BY month, p.payment\_type

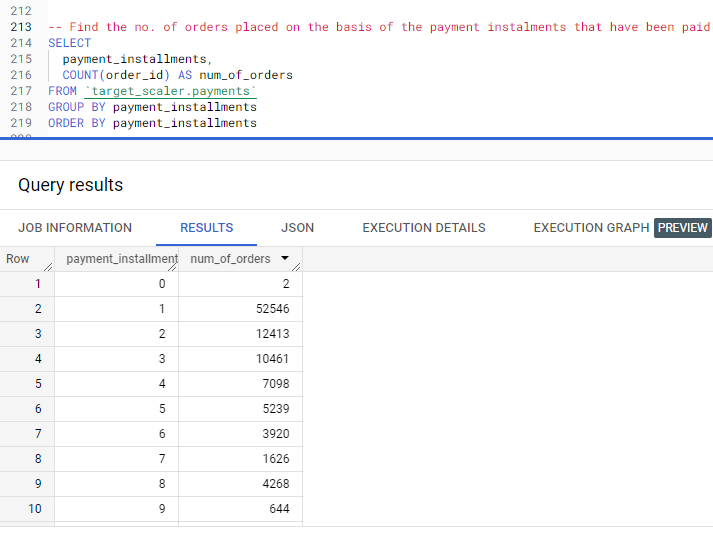
ORDER BY month, p.payment\_type, num\_of\_orders



**INSIGHTS –** We can see that majority of customers prefer credit card for payment and then UPI. Month on month the number of customers using different ways of payment has remain in same proportion. Also, as a large number of customers prefer credit card, we can tie up with different banks to offer discounts etc to attract more customer to the store and to that credit card provider as well.

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

1. SELECT
2. payment\_installments,
3. COUNT(order\_id) AS num\_of\_orders
4. FROM `target\_scaler.payments`
5. GROUP BY payment\_installments
6. ORDER BY payment\_installments



**INSIGHTS –** usually the customers are not preferring the long period of instalments. More than 50% are opting for 1 instalment only. As we saw that the highest average price is below 200, that might answer that why the majority is opting for 1 instalment only.