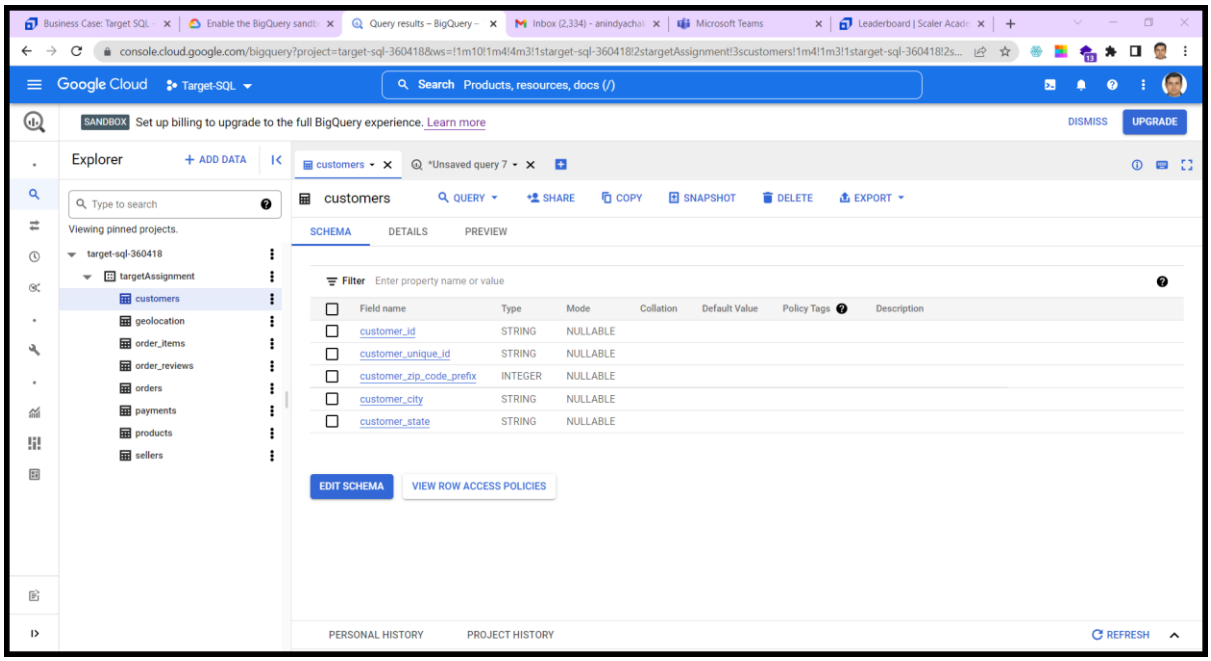
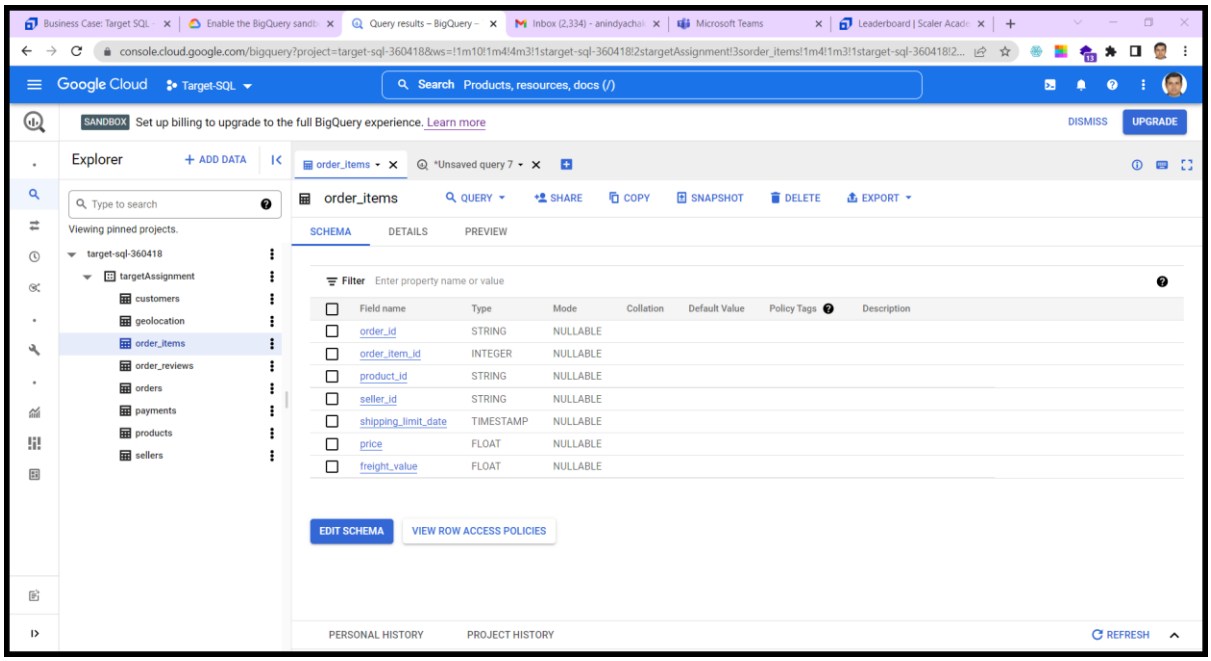
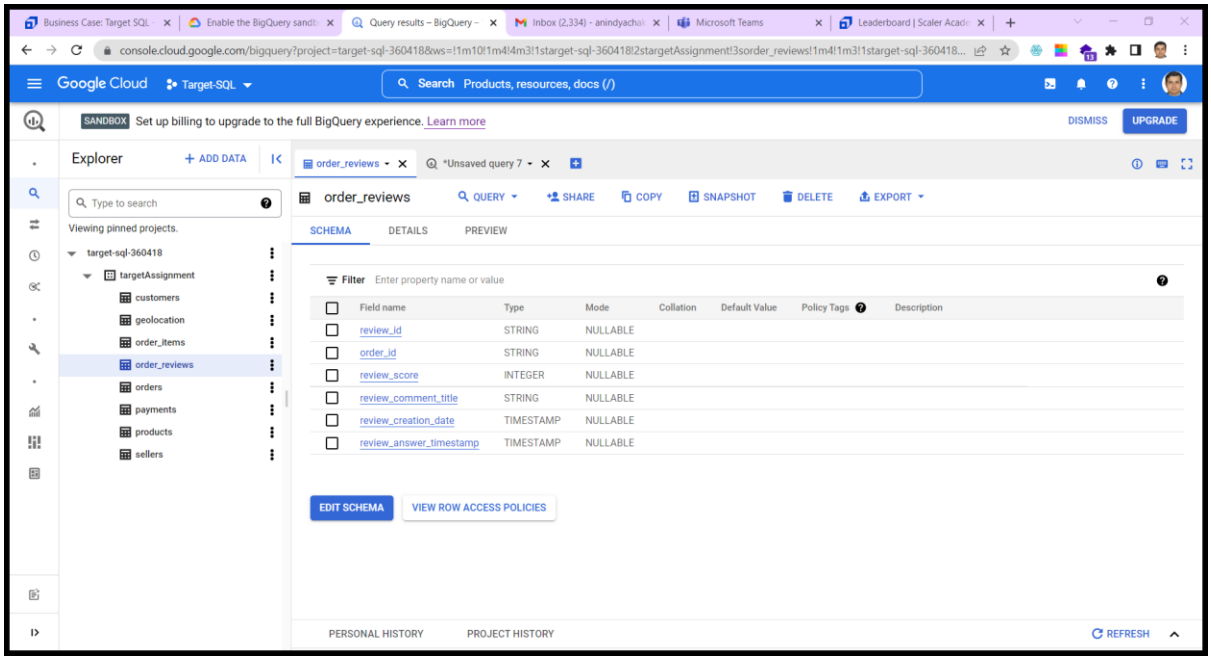
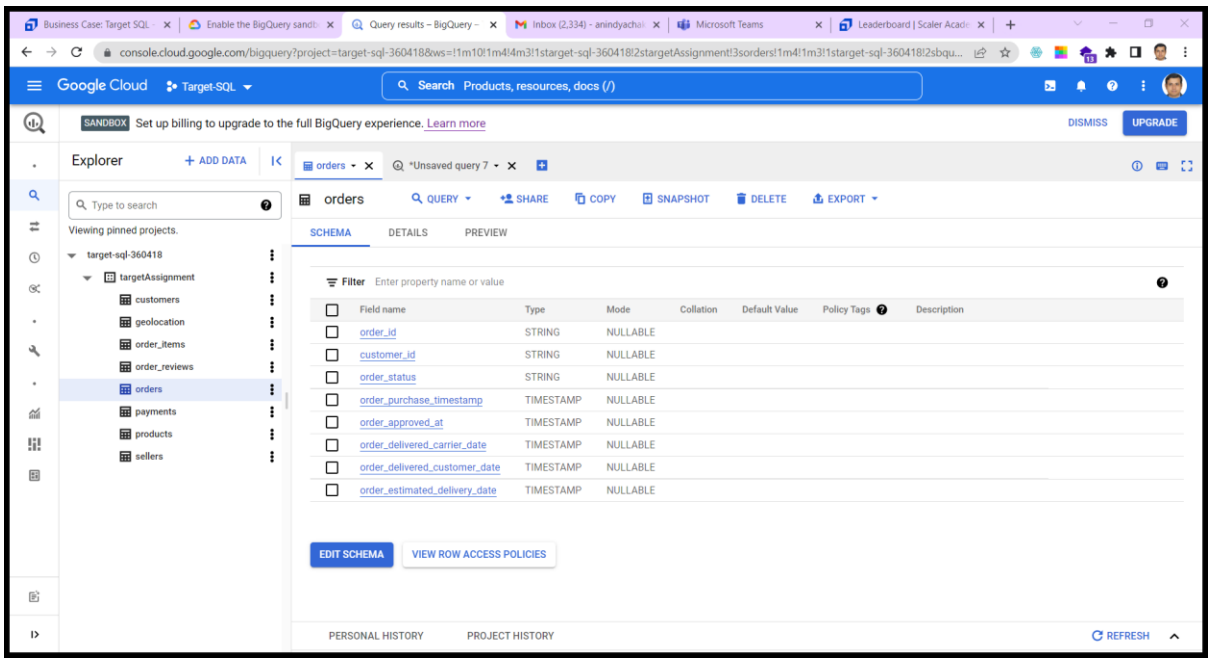
Business Case: Target SQL

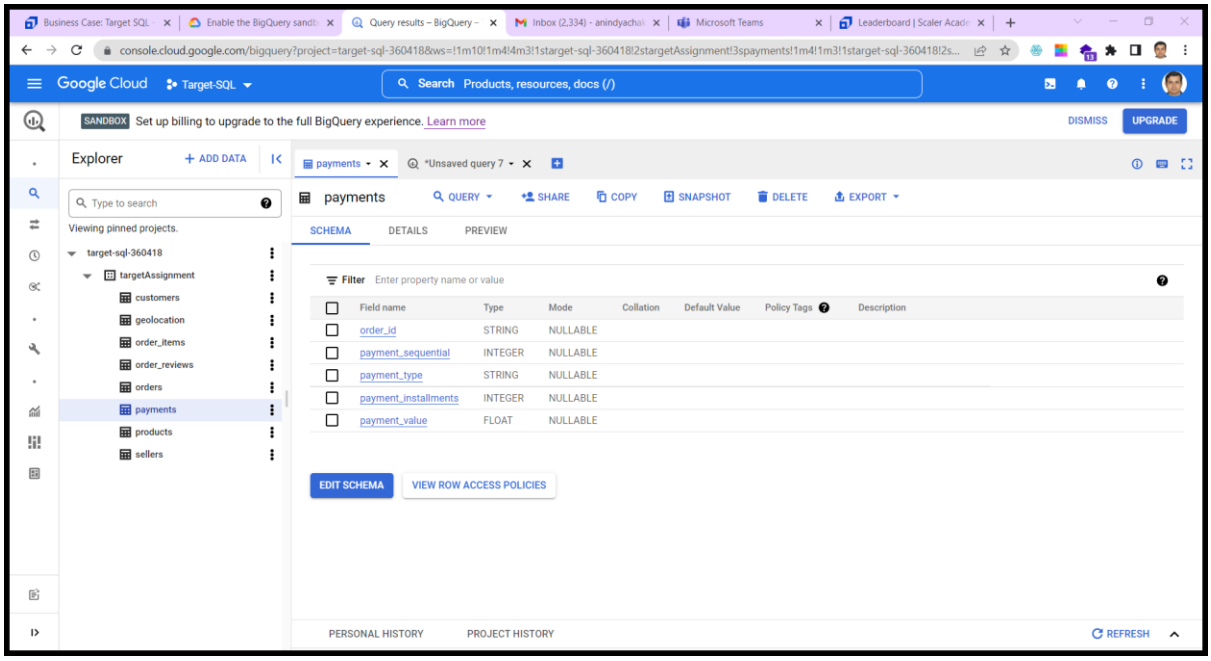
- **Initial exploration of dataset like checking the characteristics of data**
    - Import the dataset and do usual exploratory analysis steps like checking the structure & characteristics of the dataset
        - Data type of columns in a table

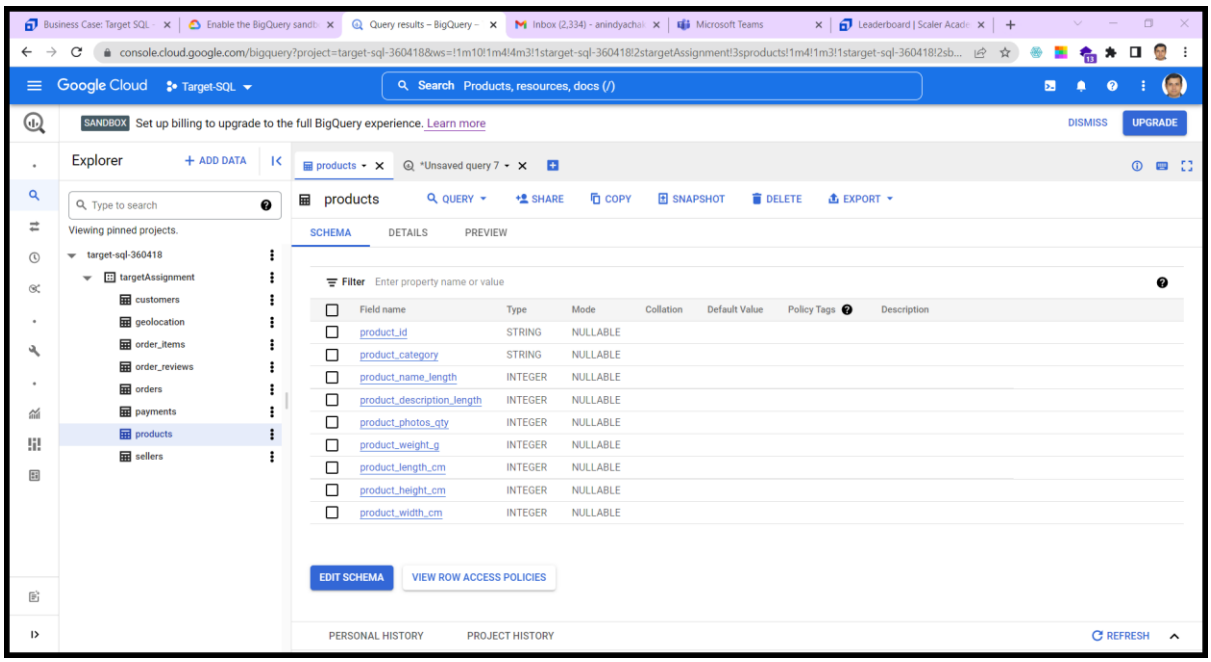


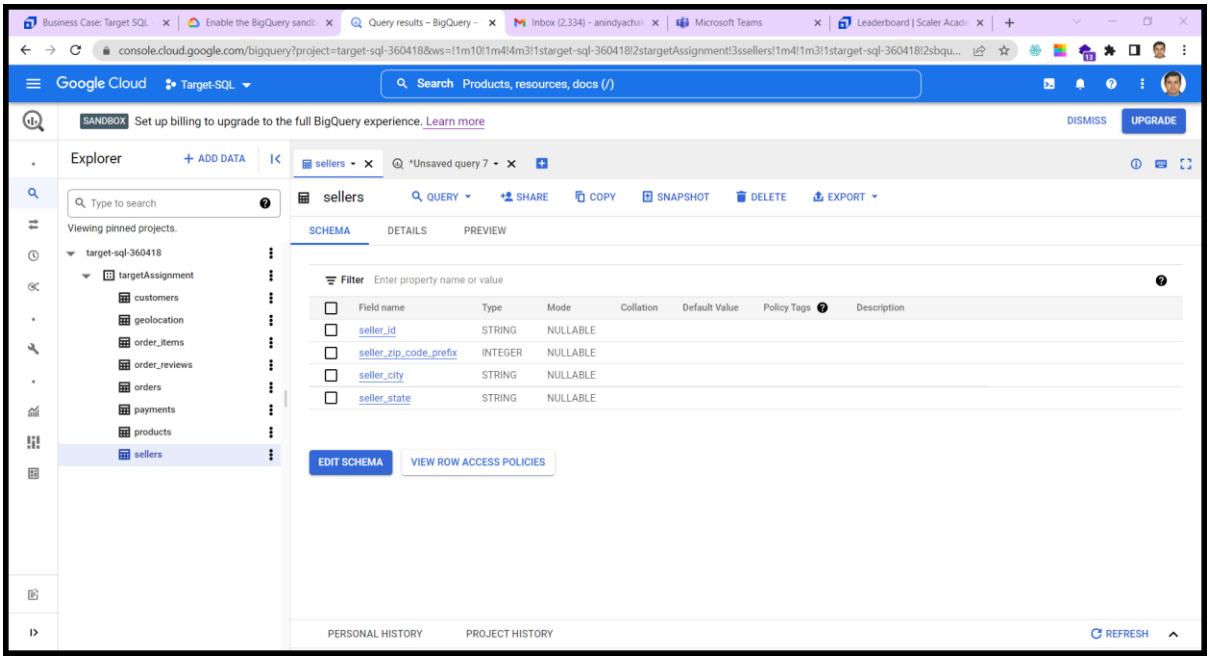


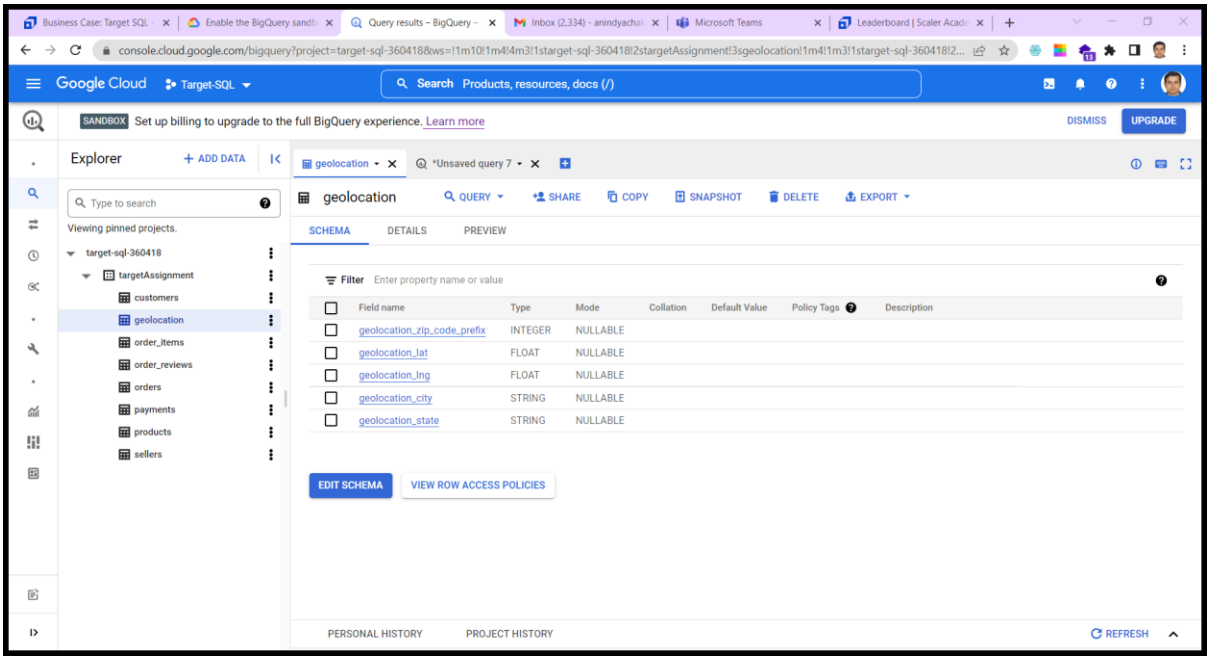

Google Cloud
Target-SQL
Search Products, resources, docs (/)
SANDBOX Set up billing to upgrade to the full BigQuery experience. Learn more
DISMISS
UPGRADE
Explorer
+ ADD DATA
Type to search
Viewing pinned projects.
target-sql-360418
targetAssignment
customers
geolocation
order_items
order_reviews
orders
payments
products
sellers
order_reviews
*Unsaved query 7
QUERY
SHARE
COPY
SNAPSHOT
DELETE
EXPORT
SCHEMA
DETAILS
PREVIEW
Filter
Enter property name or value
Field name
Type
Mode
Collation
Default Value
Policy Tags
Description
review_id STRING NULLABLE
order_id STRING NULLABLE
review_score INTEGER NULLABLE
review_comment_title STRING NULLABLE
review_creation_date TIMESTAMP NULLABLE
review_answer_timestamp TIMESTAMP NULLABLE
EDIT SCHEMA
VIEW ROW ACCESS POLICIES
PERSONAL HISTORY
PROJECT HISTORY
REFRESH


Google Cloud
Target-SQL
Search Products, resources, docs (/)
SANDBOX Set up billing to upgrade to the full BigQuery experience. Learn more
DISMISS
UPGRADE
Explorer
+ ADD DATA
Type to search
Viewing pinned projects.
target-sql-360418
targetAssignment
customers
geolocation
order_items
order_reviews
orders
payments
products
sellers
orders
*Unsaved query 7
QUERY
SHARE
COPY
SNAPSHOT
DELETE
EXPORT
SCHEMA
DETAILS
PREVIEW
Filter
Enter property name or value
Field name
Type
Mode
Collation
Default Value
Policy Tags
Description
order_id STRING NULLABLE
customer_id STRING NULLABLE
order_status STRING NULLABLE
order_purchase_timestamp TIMESTAMP NULLABLE
order_approved_at TIMESTAMP NULLABLE
order_delivered_carrier_date TIMESTAMP NULLABLE
order_delivered_customer_date TIMESTAMP NULLABLE
order_estimated_delivery_date TIMESTAMP NULLABLE
EDIT SCHEMA
VIEW ROW ACCESS POLICIES
PERSONAL HISTORY
PROJECT HISTORY
REFRESH


Google Cloud
Target-SQL
Search Products, resources, docs (/)
SANDBOX Set up billing to upgrade to the full BigQuery experience. Learn more
DISMISS
UPGRADE
Explorer
+ ADD DATA
Type to search
Viewing pinned projects.
target-sql-360418
targetAssignment
customers
geolocation
order_items
order_reviews
orders
payments
products
sellers
payments
*Unsaved query 7
QUERY
SHARE
COPY
SNAPSHOT
DELETE
EXPORT
SCHEMA
DETAILS
PREVIEW
Filter Enter property name or value
Field name Type Mode Collation Default Value Policy Tags Description
order_id STRING NULLABLE
payment_sequential INTEGER NULLABLE
payment_type STRING NULLABLE
payment_installments INTEGER NULLABLE
payment_value FLOAT NULLABLE
EDIT SCHEMA
VIEW ROW ACCESS POLICIES
PERSONAL HISTORY
PROJECT HISTORY
REFRESH

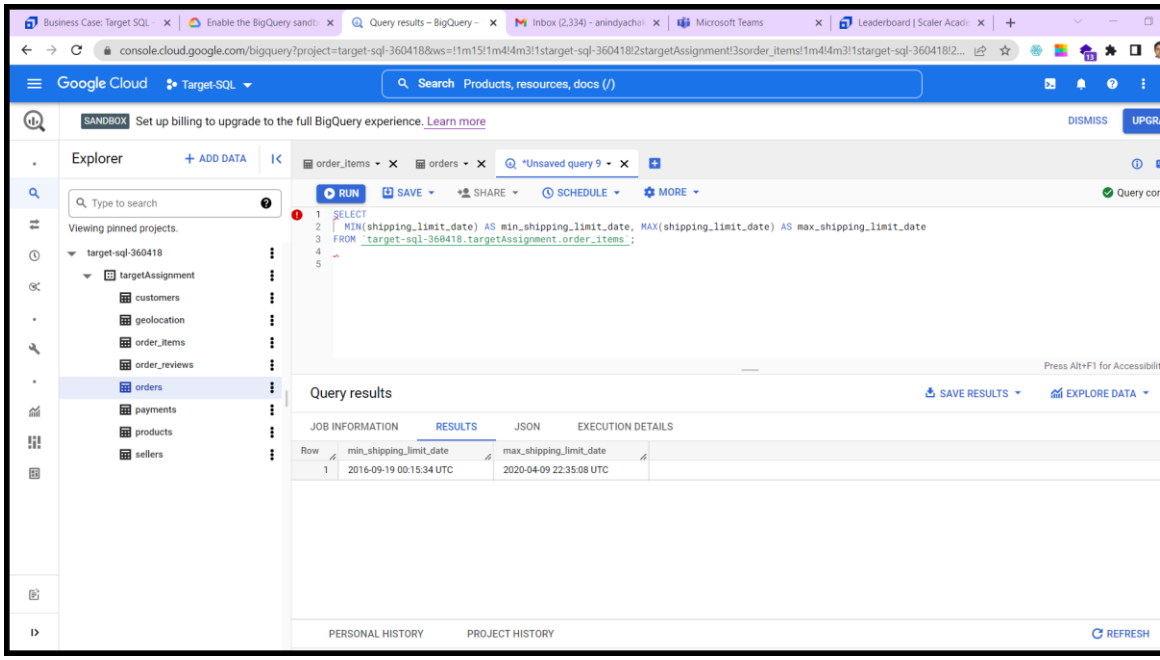

Google Cloud
Target-SQL
Search Products, resources, docs (/)
SANDBOX Set up billing to upgrade to the full BigQuery experience. Learn more
DISMISS
UPGRADE
Explorer
+ ADD DATA
Type to search
Viewing pinned projects.
target-sql-360418
targetAssignment
customers
geolocation
order_items
order_reviews
orders
payments
products
sellers
products
*Unsaved query 7
QUERY
SHARE
COPY
SNAPSHOT
DELETE
EXPORT
SCHEMA
DETAILS
PREVIEW
Filter Enter property name or value
Field name Type Mode Collation Default Value Policy Tags Description
product_id STRING NULLABLE
product_category STRING NULLABLE
product_name_length INTEGER NULLABLE
product_description_length INTEGER NULLABLE
product_photos_qty INTEGER NULLABLE
product_weight_g INTEGER NULLABLE
product_length_cm INTEGER NULLABLE
product_height_cm INTEGER NULLABLE
product_width_cm INTEGER NULLABLE
EDIT SCHEMA
VIEW ROW ACCESS POLICIES
PERSONAL HISTORY
PROJECT HISTORY
REFRESH

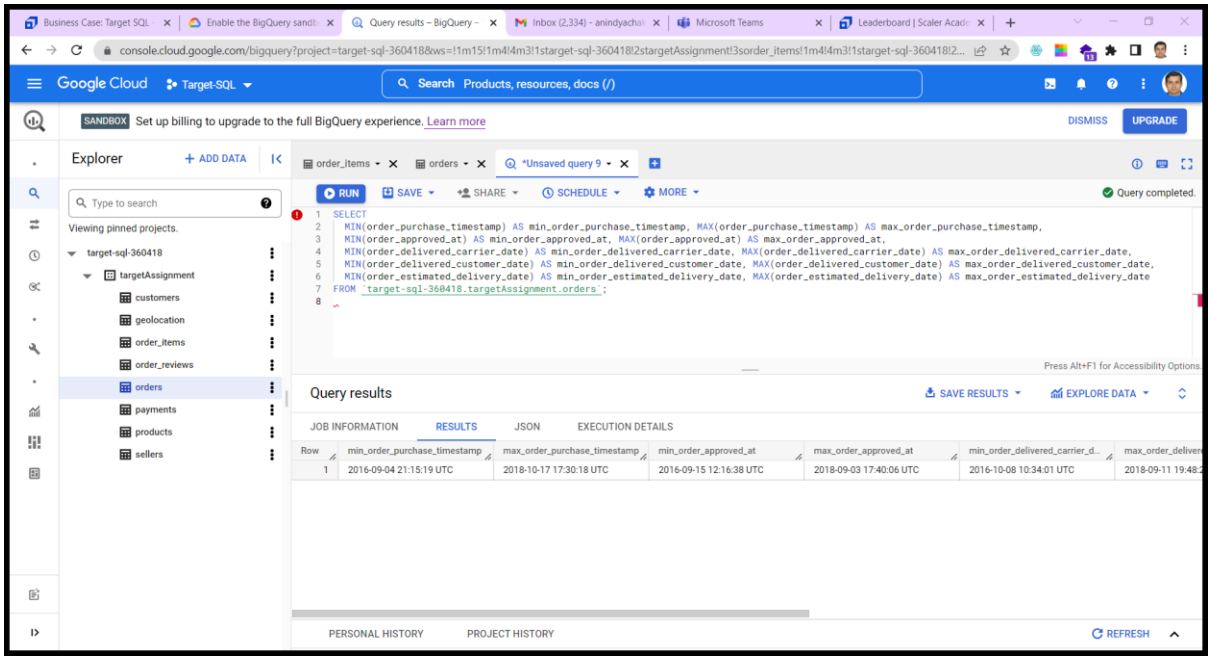




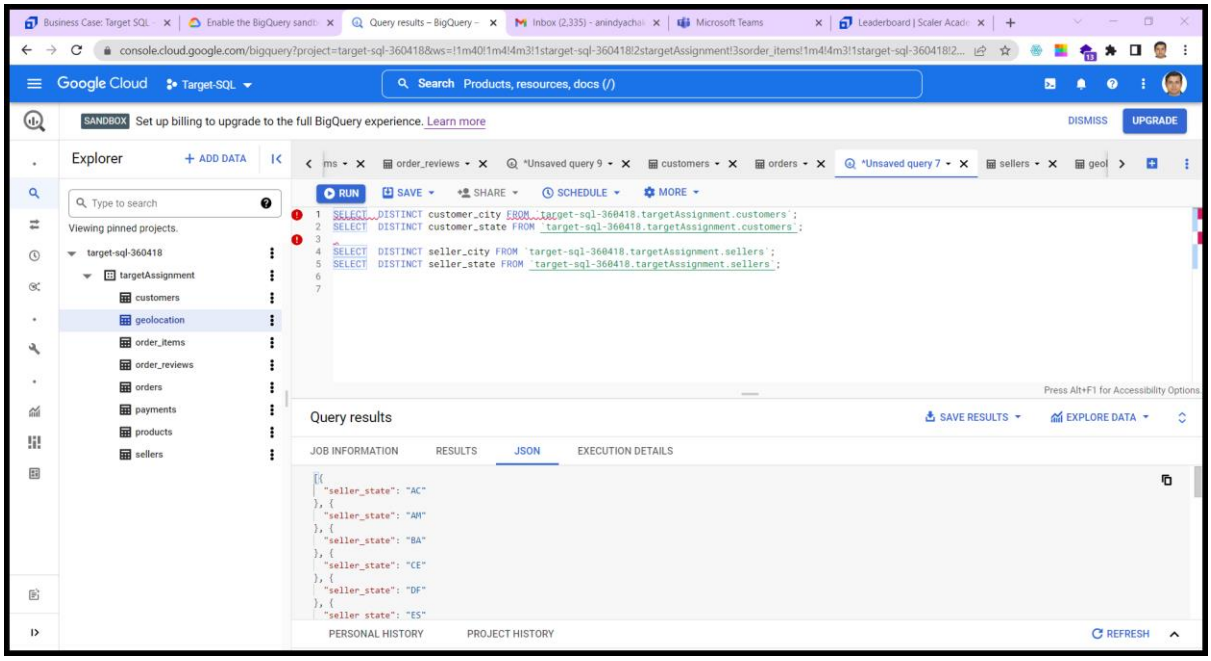
- Time period for which the data is given

  - ✓ **Orders** – From 2016-09-04 21:15:19 UTC to 2018-11-12 00:00:00 UTC
  - ✓ **Order_reviews** - From 2016-09-19 00:15:34 UTC to 2020-04-09 22:35:08 UTC



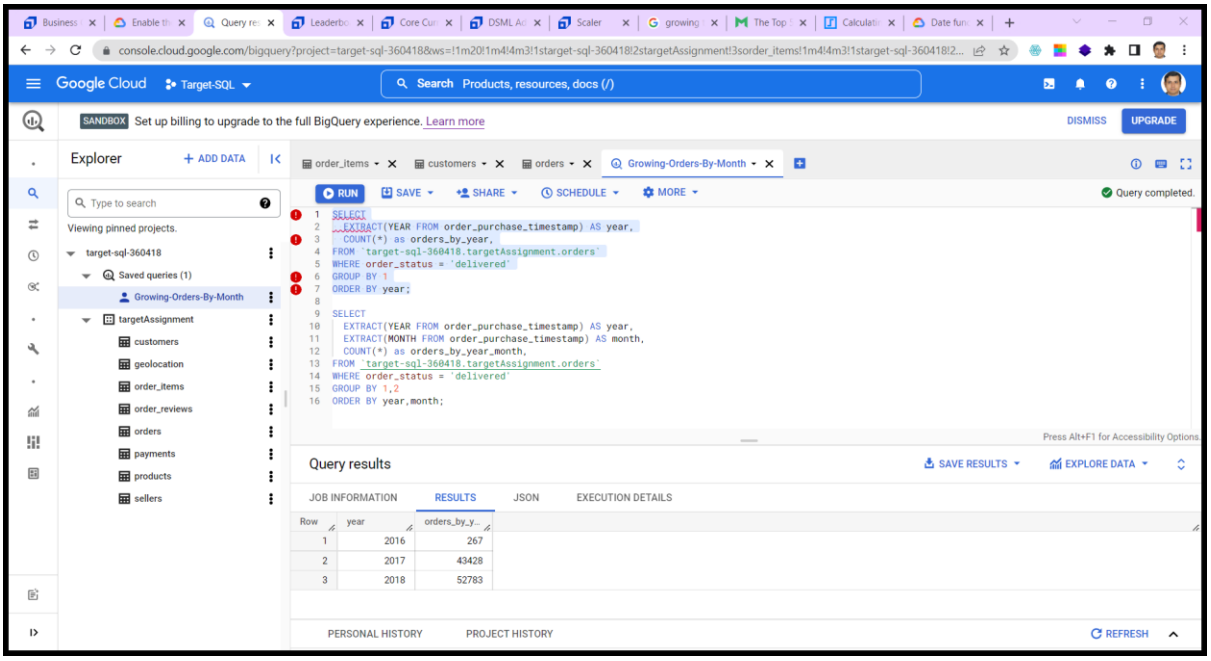


- Cities and States covered in the dataset



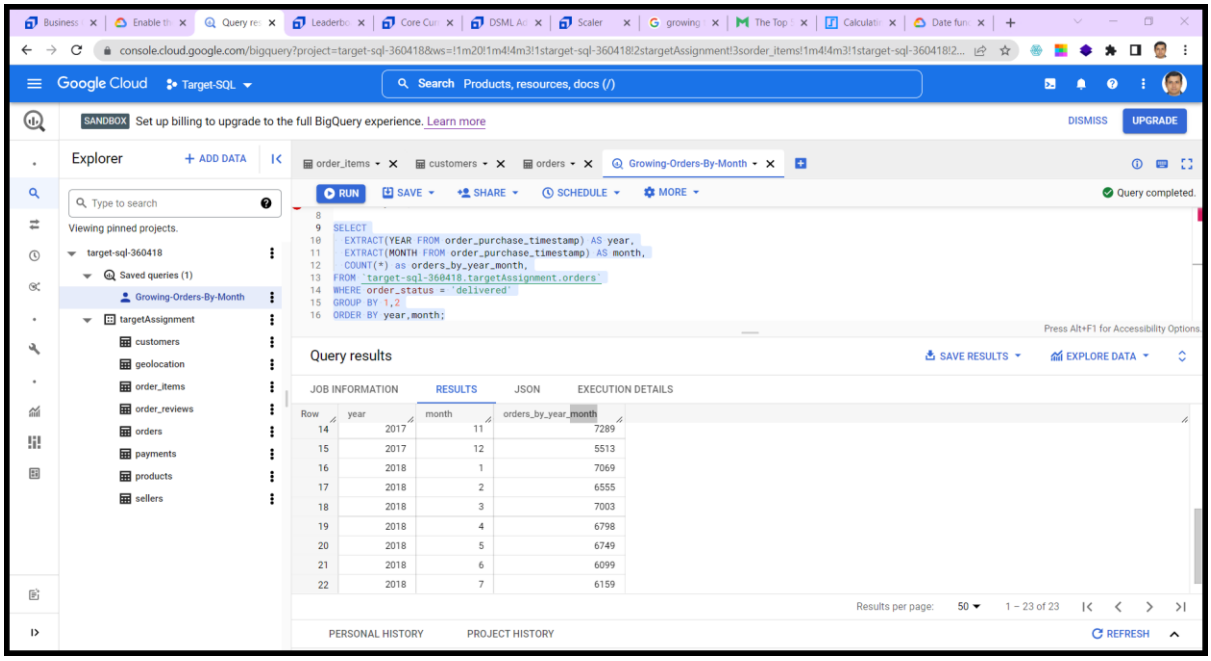
- **In-depth Exploration**

  - Is there a growing trend on e-commerce in Brazil?
    - Yes, there is a growing trend year on year since 2017 onwards.

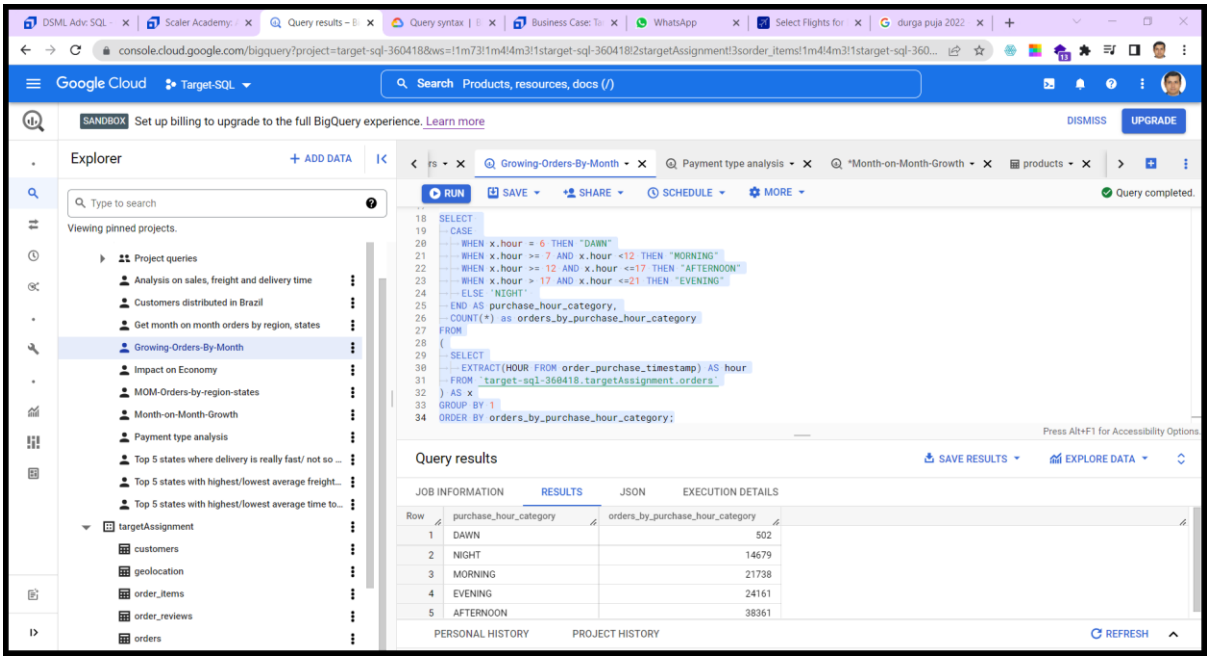


  - How can we describe a complete scenario?
    - In year 2016, it was just the beginning it seems
    - In year 2017, it started showing **almost growing** trends every month on month, till month of November
    - There was slight decline in month of December 2017

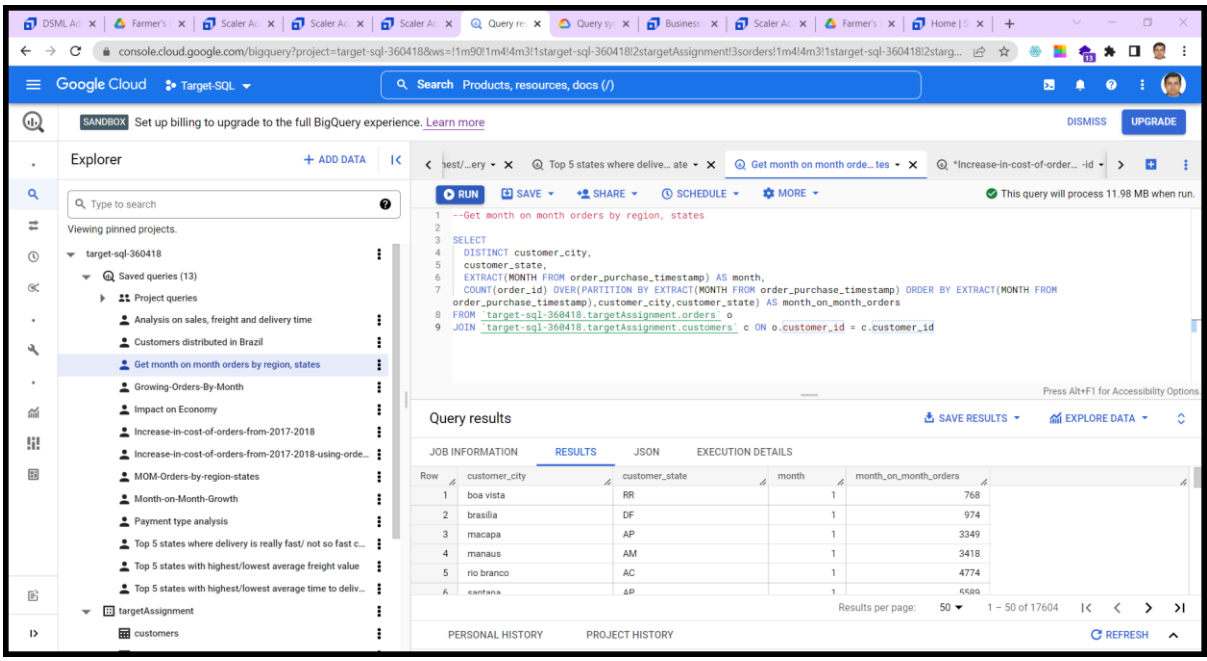
- In year 2018 January onwards, it showed again consistent orders of 6K + per month for the rest of year



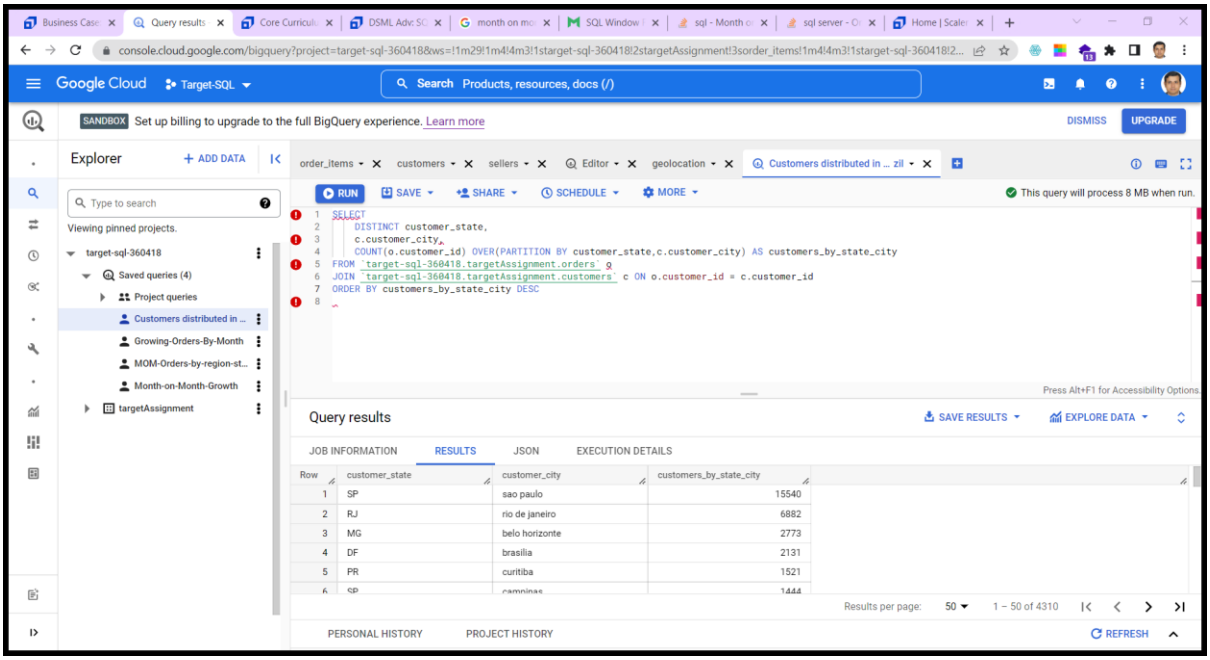
- Can we see some seasonality with peaks at specific months?
  - More orders in months of January, March, May etc. i.e., in the first half of the year
  - In later half of the year i.e. in months of July to August (based on data availability in year 2018) , there are consistent good volume of orders
- What time do Brazilian customers tend to buy (Dawn, Morning, Afternoon or Night)?
  - Most of the purchases are during Morning, Evening and Afternoon.
  - There are few purchases in Early morning as well i.e., between 0^{th} hour and 5^{th} Hour



- **Evolution of E-commerce orders in the Brazil region**
    - Get month on month orders by region, states

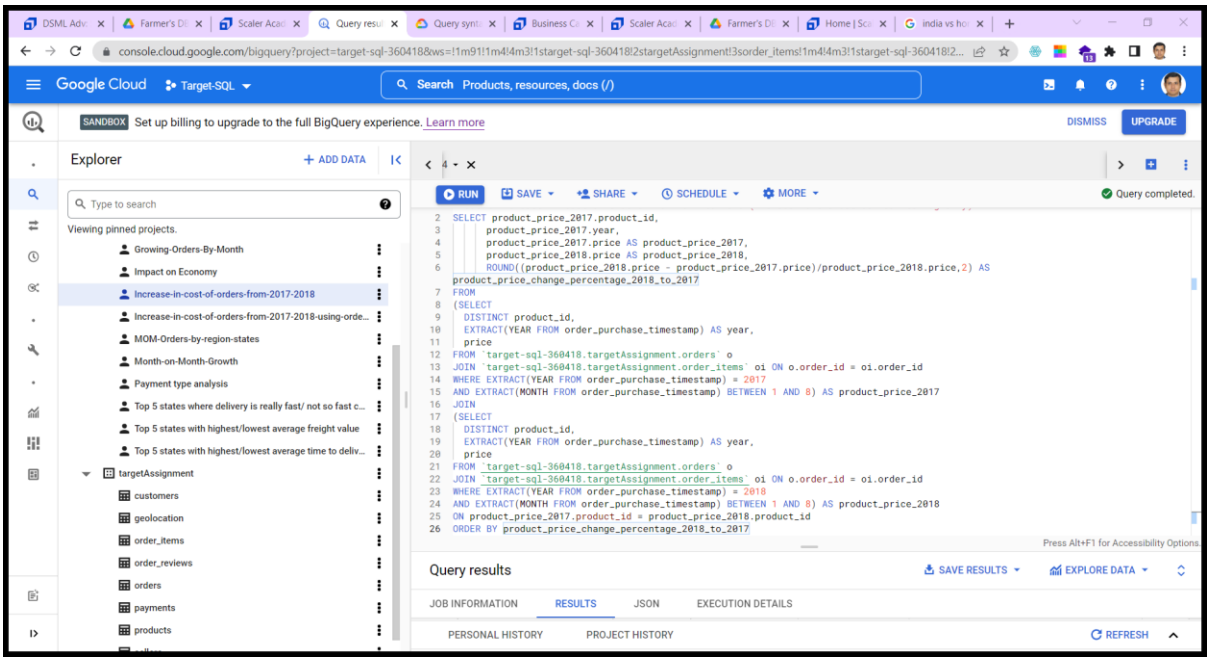


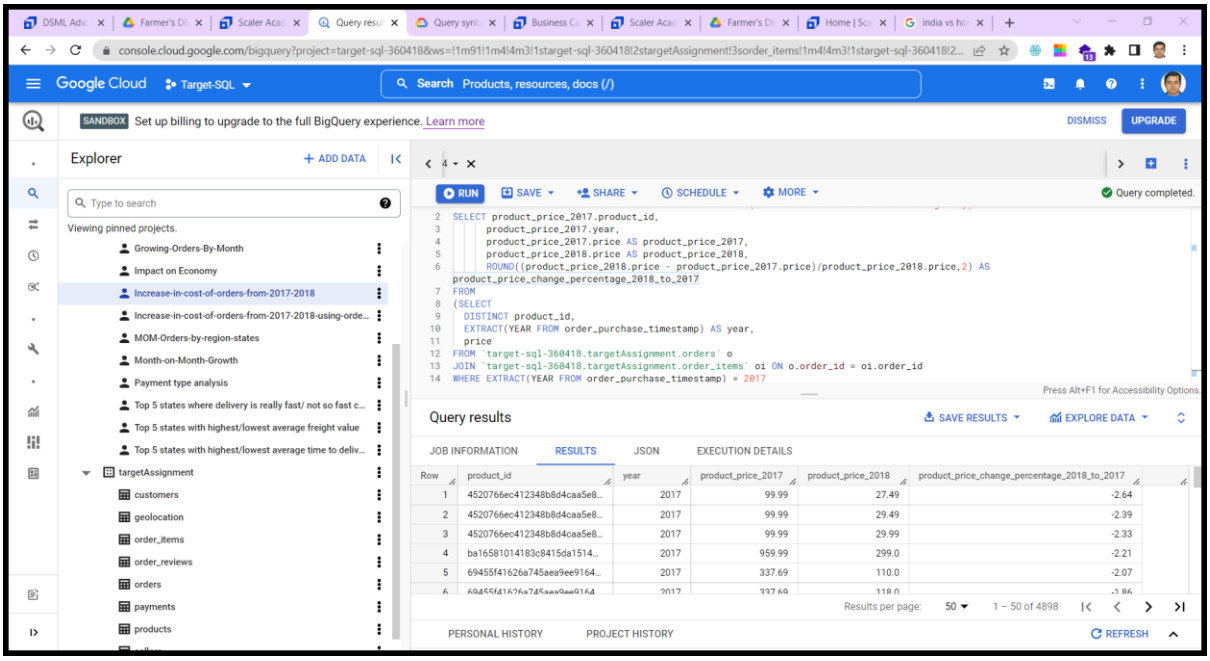
    - How are customers distributed in Brazil
        - Major customers are from Sao Paulo, Rio De Janeiro, Belo Horizonte



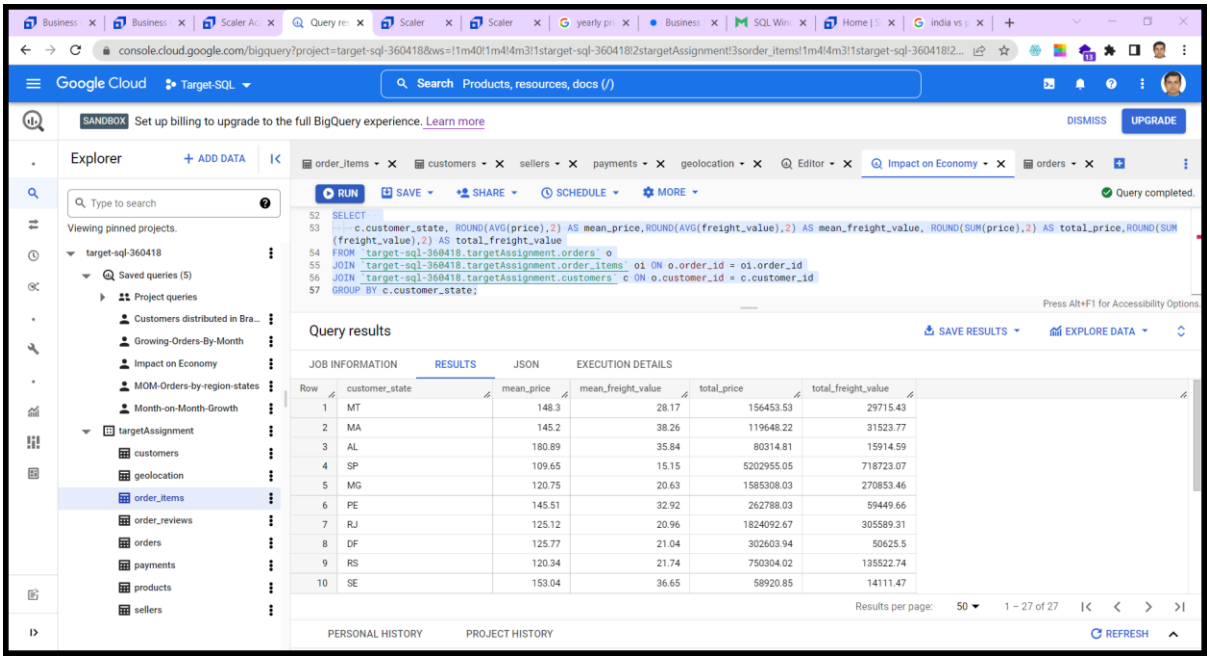
- **Impact on Economy - Analyse the money movement by e-commerce by looking at order prices, freight and others.**

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



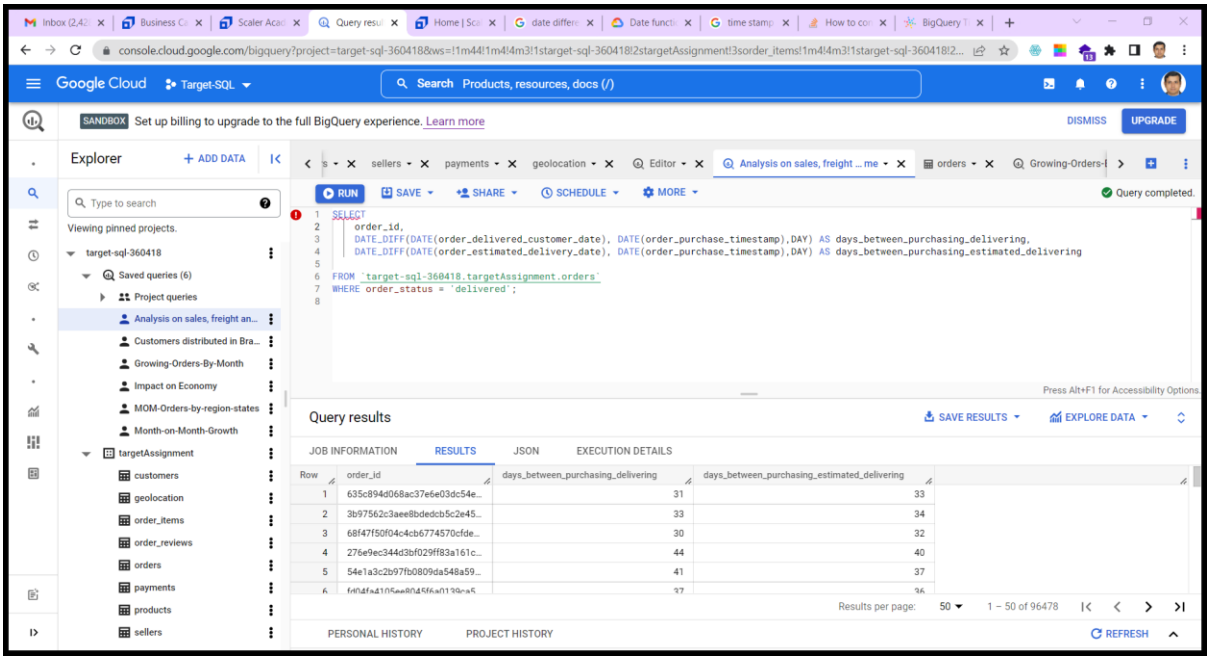


- Mean & Sum of price and freight value by customer state

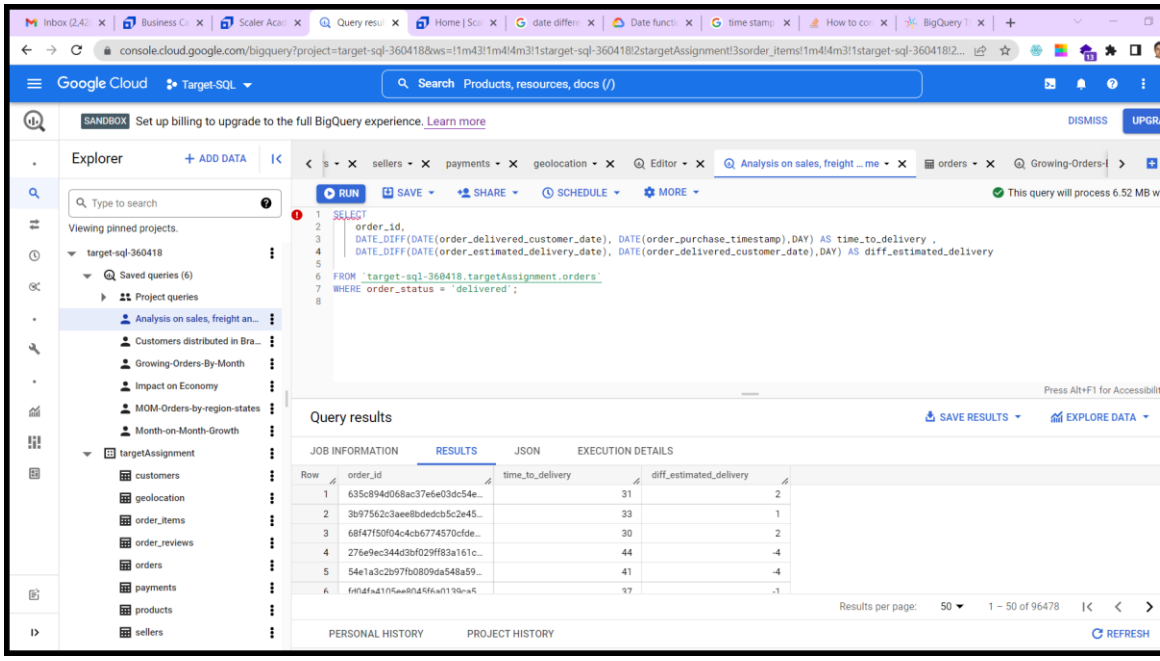


- **Analysis on sales, freight and delivery time**

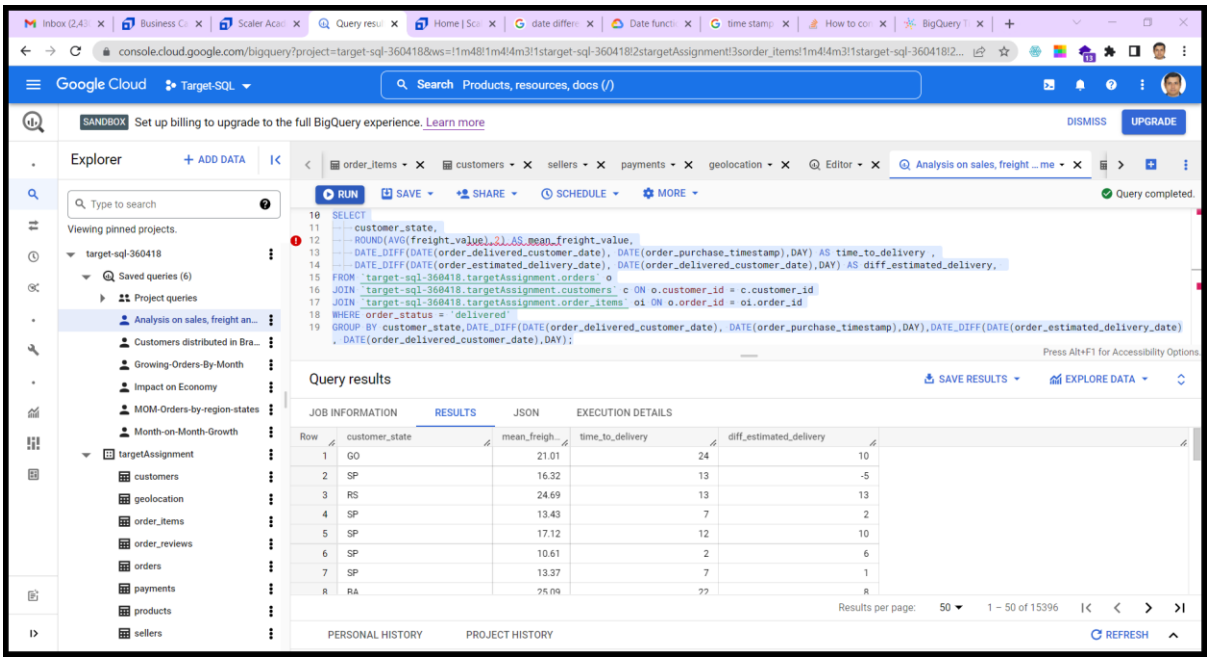
  - Calculate days between purchasing, delivering and estimated delivery



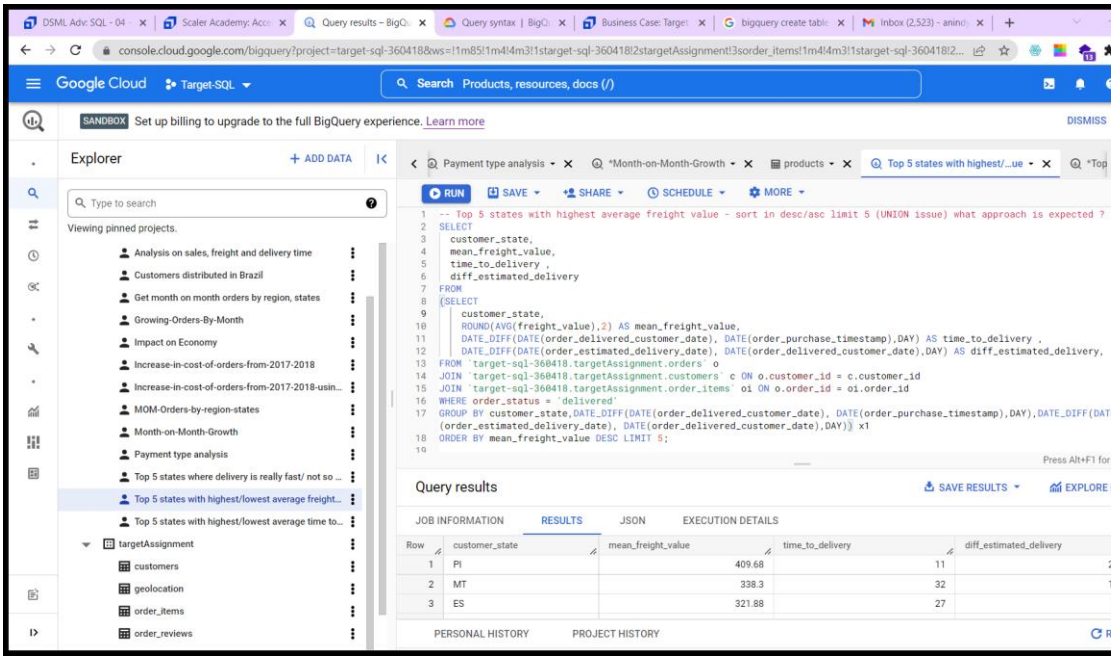
- Create columns:
    - time_to_delivery = order_purchase_timestamp-order_delivered_customer_date
    - diff_estimated_delivery = order_estimated_delivery_date-order_delivered_customer_date

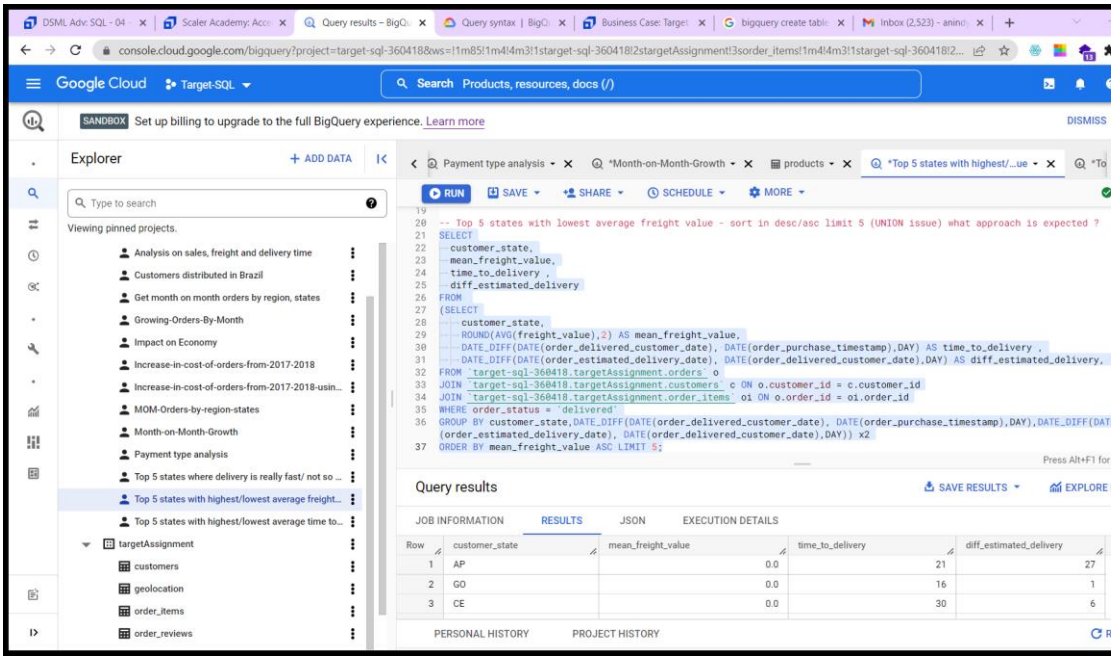


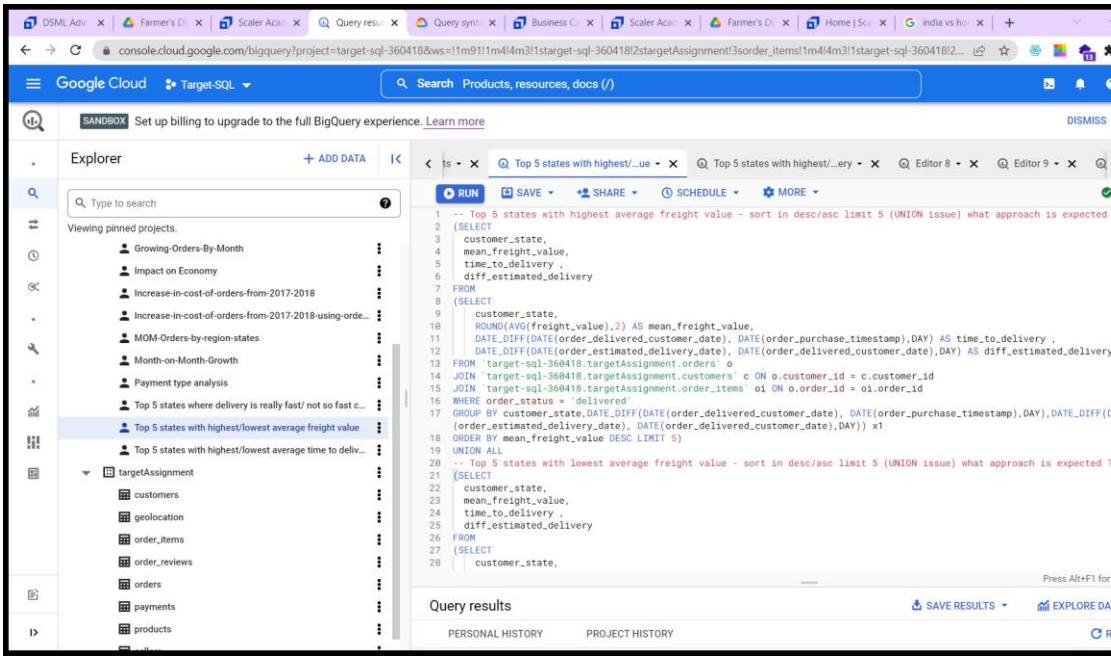
- Group data by state, take mean of freight_value, time_to_delivery, diff_estimated_delivery



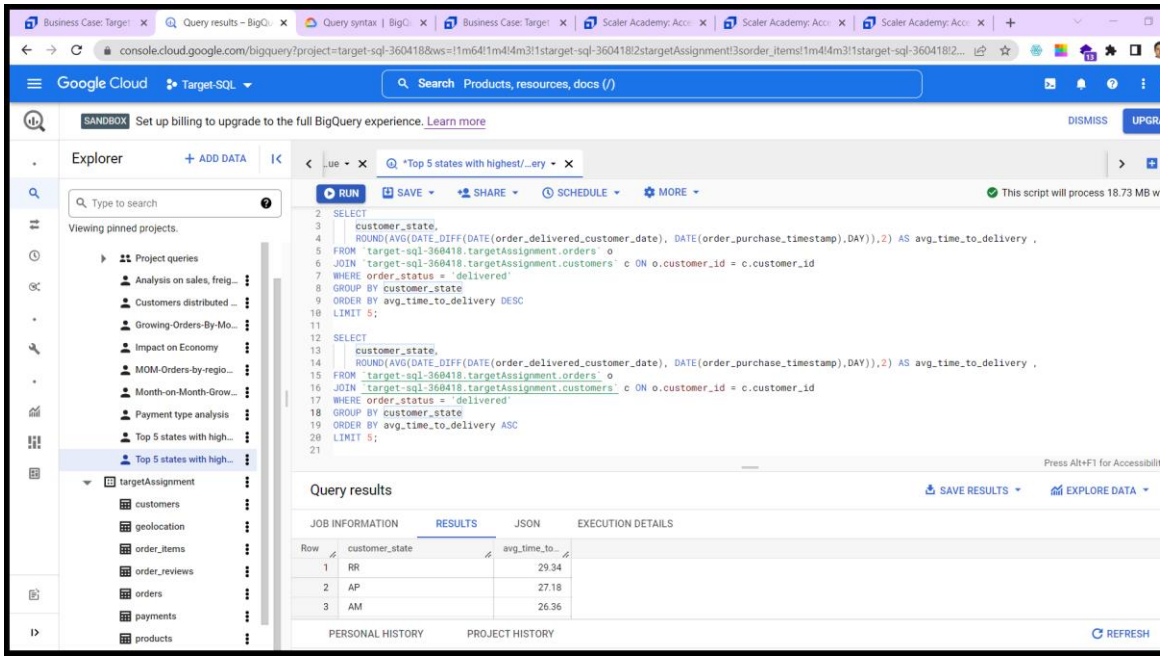
- Sort the data to get the following:
  - Top 5 states with highest/lowest average freight value - sort in desc/asc limit 5?

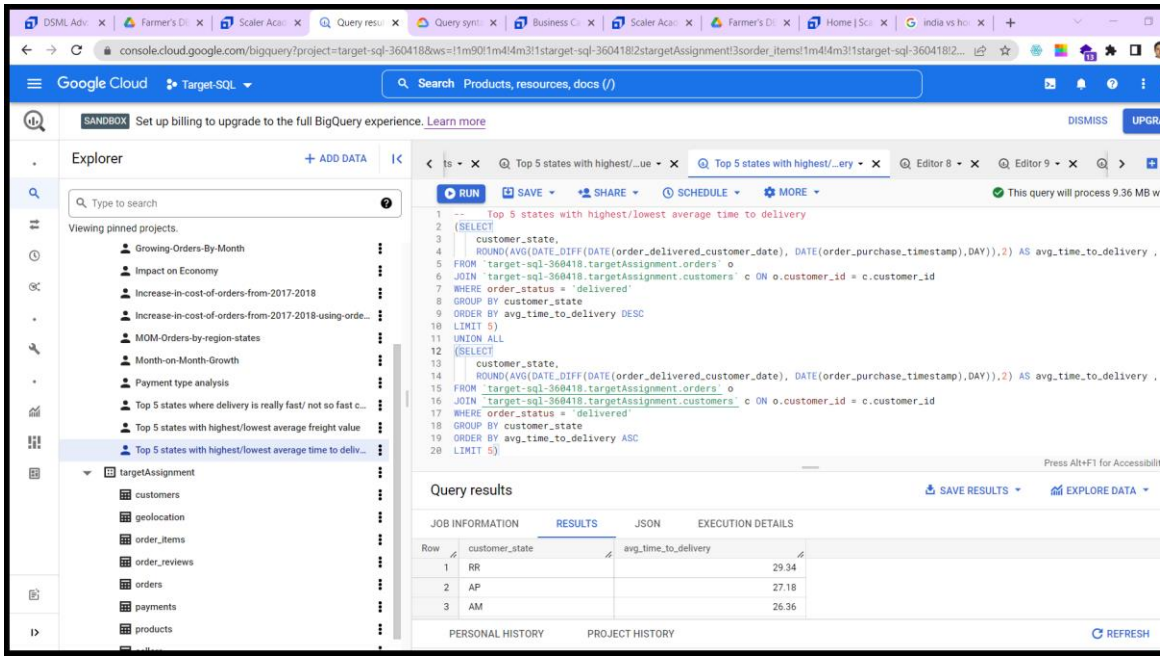

Google Cloud
Target-SQL
Search Products, resources, docs (/)
SANDBOX Set up billing to upgrade to the full BigQuery experience. Learn more
DISMISS
Explorer
+ ADD DATA
Type to search
Viewing pinned projects.
Analysis on sales, freight and delivery time
Customers distributed in Brazil
Get month on month orders by region, states
Growing-Orders-By-Month
Impact on Economy
Increase-in-cost-of-orders-from-2017-2018
Increase-in-cost-of-orders-from-2017-2018-usin...
MOM-Orders-by-region-states
Month-on-Month-Growth
Payment type analysis
Top 5 states where delivery is really fast/ not so ...
Top 5 states with highest/lowest average freight...
Top 5 states with highest/lowest average time to...
targetAssignment
customers
geolocation
order_items
order_reviews
Payment type analysis
*Month-on-Month-Growth
products
Top 5 states with highest/...ue
RUN
SAVE
SHARE
SCHEDULE
MORE
-- Top 5 states with highest average freight value - sort in desc/asc limit 5 (UNION issue) what approach is expected ?
SELECT
customer_state,
mean_freight_value,
time_to_delivery ,
diff_estimated_delivery
FROM
(SELECT
customer_state,
ROUND(AVG(freight_value),2) AS mean_freight_value,
DATE_DIFF(DATE(order_delivered_customer_date), DATE(order_purchase_timestamp),DAY) AS time_to_delivery ,
DATE_DIFF(DATE(order_estimated_delivery_date), DATE(order_delivered_customer_date),DAY) AS diff_estimated_delivery,
FROM `target-sql-360418.targetAssignment.orders` o
JOIN `target-sql-360418.targetAssignment.customers` c ON o.customer_id = c.customer_id
JOIN `target-sql-360418.targetAssignment.order_items` oi ON o.order_id = oi.order_id
WHERE order_status = 'delivered'
GROUP BY customer_state,DATE_DIFF(DATE(order_delivered_customer_date), DATE(order_purchase_timestamp),DAY),DATE_DIFF(DAT
(order_estimated_delivery_date), DATE(order_delivered_customer_date),DAY)) x1
ORDER BY mean_freight_value DESC LIMIT 5;
Query results
SAVE RESULTS
EXPLORE
JOB INFORMATION
RESULTS
JSON
EXECUTION DETAILS
Row customer_state mean_freight_value time_to_delivery diff_estimated_delivery
1 PI 409.68 11
2 MT 338.3 32
3 ES 321.88 27
PERSONAL HISTORY
PROJECT HISTORY


Google Cloud
Target-SQL
Search Products, resources, docs (/)
SANDBOX Set up billing to upgrade to the full BigQuery experience. Learn more
DISMISS
Explorer
+ ADD DATA
Type to search
Viewing pinned projects.
Analysis on sales, freight and delivery time
Customers distributed in Brazil
Get month on month orders by region, states
Growing-Orders-By-Month
Impact on Economy
Increase-in-cost-of-orders-from-2017-2018
Increase-in-cost-of-orders-from-2017-2018-usin...
MOM-Orders-by-region-states
Month-on-Month-Growth
Payment type analysis
Top 5 states where delivery is really fast/ not so ...
Top 5 states with highest/lowest average freight...
Top 5 states with highest/lowest average time to...
targetAssignment
customers
geolocation
order_items
order_reviews
Payment type analysis
*Month-on-Month-Growth
products
*Top 5 states with highest/...ue
RUN
SAVE
SHARE
SCHEDULE
MORE
-- Top 5 states with lowest average freight value - sort in desc/asc limit 5 (UNION issue) what approach is expected ?
SELECT
customer_state,
mean_freight_value,
time_to_delivery ,
diff_estimated_delivery
FROM
(SELECT
customer_state,
ROUND(AVG(freight_value),2) AS mean_freight_value,
DATE_DIFF(DATE(order_delivered_customer_date), DATE(order_purchase_timestamp),DAY) AS time_to_delivery ,
DATE_DIFF(DATE(order_estimated_delivery_date), DATE(order_delivered_customer_date),DAY) AS diff_estimated_delivery,
FROM `target-sql-360418.targetAssignment.orders` o
JOIN `target-sql-360418.targetAssignment.customers` c ON o.customer_id = c.customer_id
JOIN `target-sql-360418.targetAssignment.order_items` oi ON o.order_id = oi.order_id
WHERE order_status = 'delivered'
GROUP BY customer_state,DATE_DIFF(DATE(order_delivered_customer_date), DATE(order_purchase_timestamp),DAY),DATE_DIFF(DAT
(order_estimated_delivery_date), DATE(order_delivered_customer_date),DAY)) x2
ORDER BY mean_freight_value ASC LIMIT 5;
Query results
SAVE RESULTS
EXPLORE
JOB INFORMATION
RESULTS
JSON
EXECUTION DETAILS
Row customer_state mean_freight_value time_to_delivery diff_estimated_delivery
1 AP 0.0 21 27
2 GO 0.0 16 1
3 CE 0.0 30 6
PERSONAL HISTORY
PROJECT HISTORY

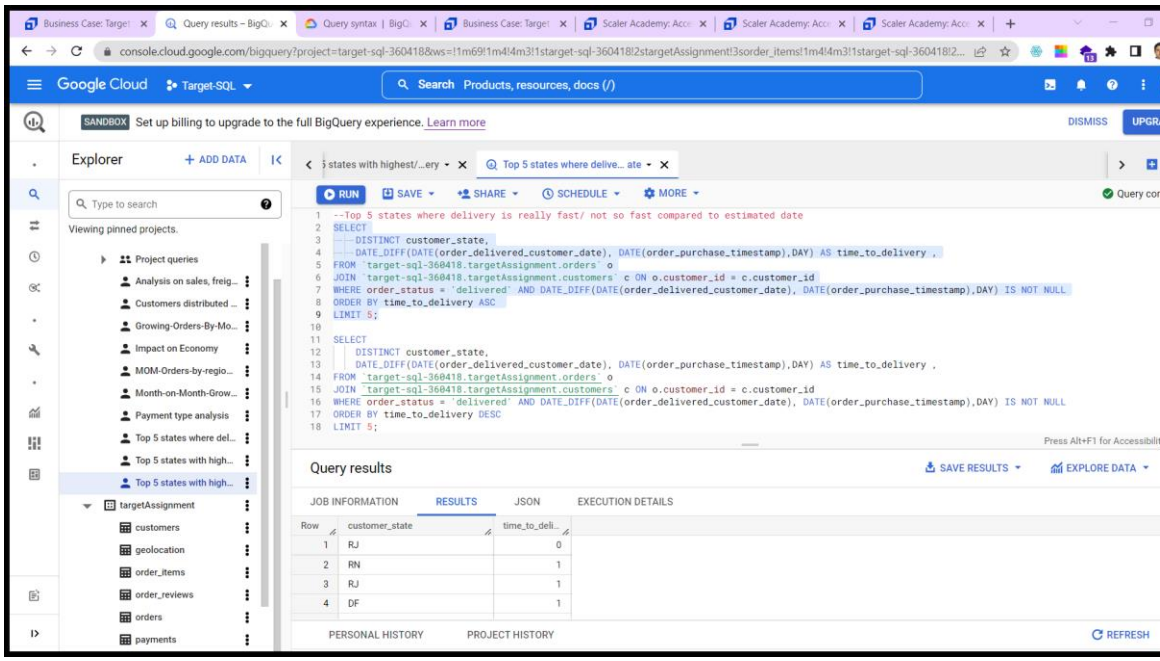


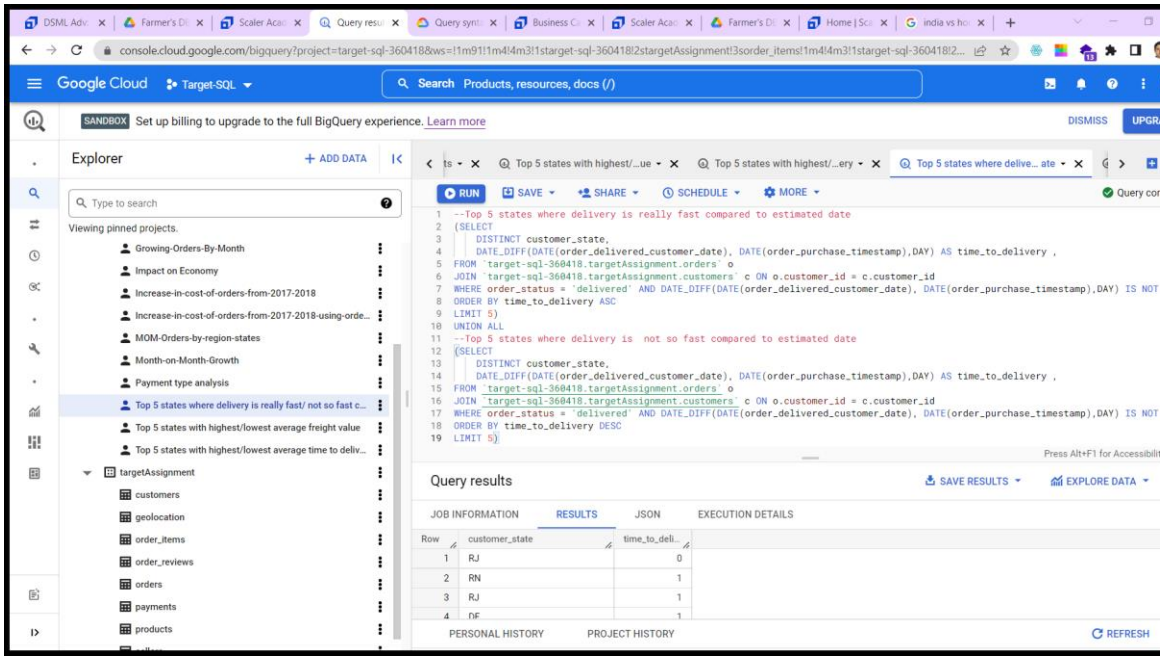
- Top 5 states with highest/lowest average time to delivery



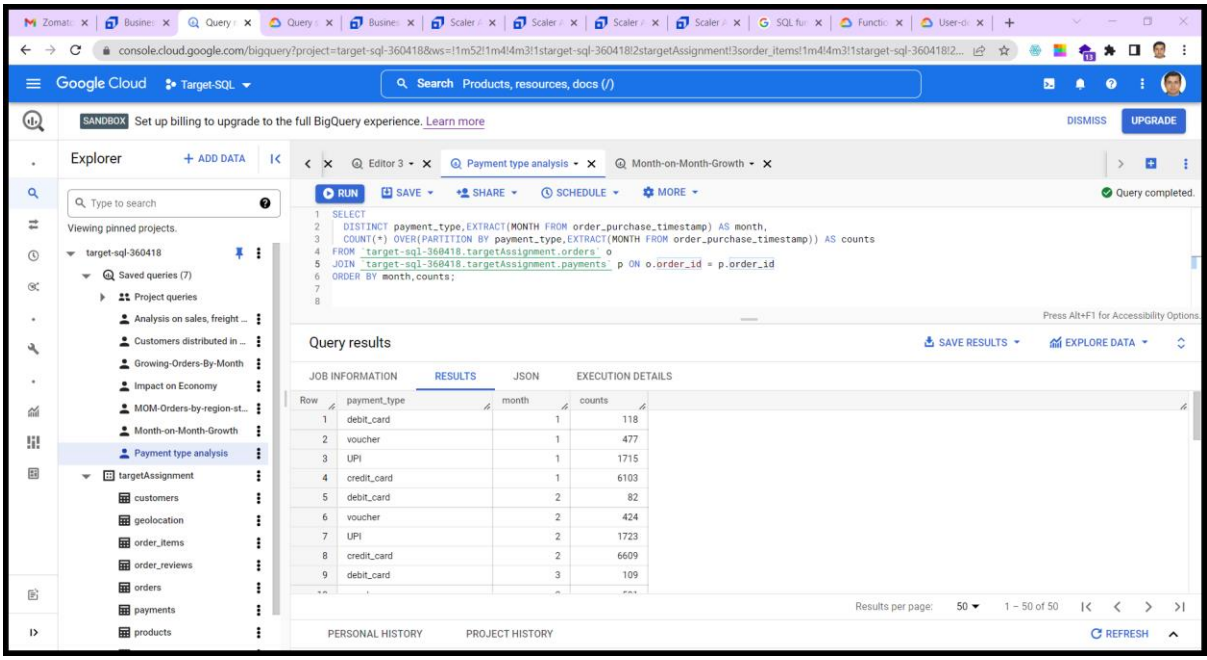


- Top 5 states where delivery is really fast/ not so fast compared to estimated date

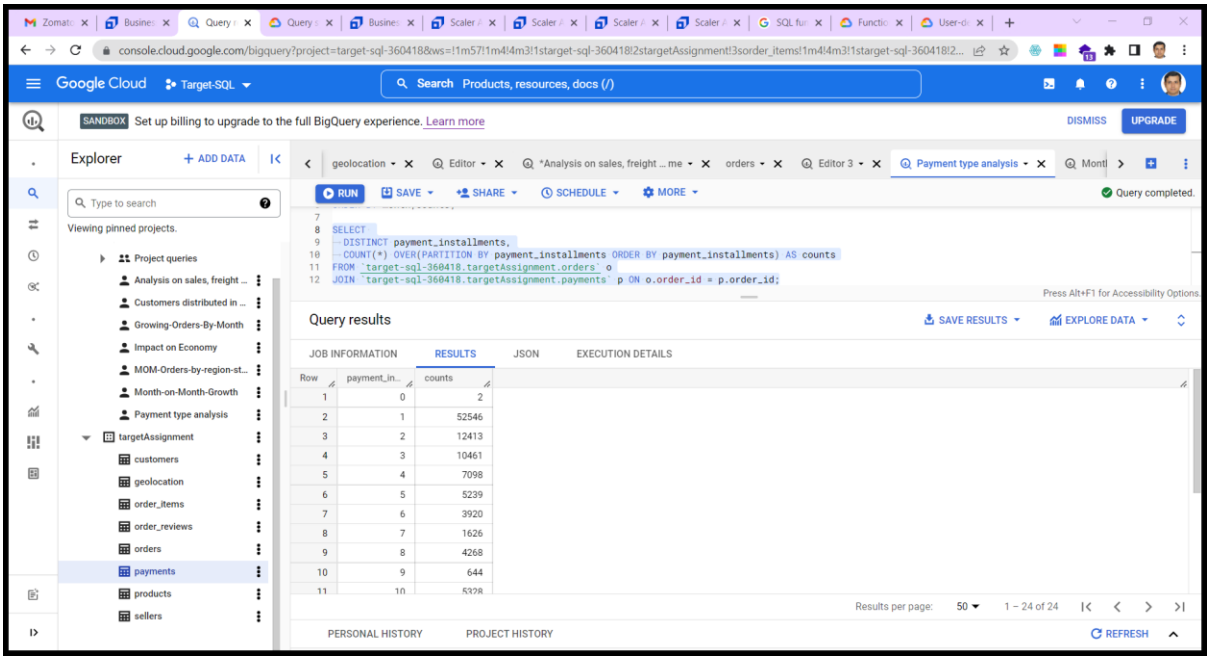




- **Payment type analysis**
    - Month over Month count of orders for different payment types



    - Distribution of payment instalments and count of orders



- **Actionable Insights**
  - There is a growing trend of e-commerce from 2016 to 2018
  - There is consistent order volume across the months
  - Highest number of orders placed during afternoon (highest), Evening and morning (in the order of sequence)
  - There are few orders placed during Dawn, it will be interesting to see if orders placed during Dawn are from specific customers, region or for specific product categories
  - Major orders are placed from city Sao Paulo, Rio De Janeiro, Belo Horizonte
  - Actual delivery is faster than estimated time of delivery
  - TOP 5 states which has highest freight values, have time to delivery of max 33 days
  - RR, AP, AM, AL and PA are top 5 states where average delivery time is highest
  - SP, PR, MG, DF, SC are top 5 states where average delivery time is lowest
  - Delivery time is fastest in states such as RJ, RN, DF
  - Delivery time is slowest in states such as ES, RJ, PA, PI etc.
  - Most of the payments are via credit card (highest), UPI and voucher(lowest) (in order of their sequence)
- **Recommendations**
  - There should more promotions in the regions of fastest delivery with committed delivery time to raise sales (e.g., states SP, PR, MG, DF, SC etc.)
  - More promotions during Afternoon and Evening as there are most of the orders being placed or early morning sales for customers to boost sales during off hours
  - Season based promotions during festive season such December, January or in November (during Nation Holidays) or in other regional Carnival days
  - Special promotions during Father's Day (in August), mothers' day (in May)

- Product based marketing can be another option based on customer profile or customer geo locations.
- Wallet based cashbacks can be offered for customers to minimize credit /UPI transaction costs.
- As there is a growing trend, social media marketing channels can be targeted or referral system can be introduced to avail discounts on purchases.