**Business Case: Target SQL**

**Context:**

**Target is a globally renowned brand and a prominent retailer in the United States. Target makes itself a preferred shopping destination by offering outstanding value, inspiration, innovation and an exceptional guest experience that no other retailer can deliver.**

**This particular business case focuses on the operations of Target in Brazil and provides insightful information about 100,000 orders placed between 2016 and 2018. The dataset offers a comprehensive view of various dimensions including the order status, price, payment and freight performance, customer location, product attributes, and customer reviews.**

**By analyzing this extensive dataset, it becomes possible to gain valuable insights into Target's operations in Brazil. The information can shed light on various aspects of the business, such as order processing, pricing strategies, payment and shipping efficiency, customer demographics, product characteristics, and customer satisfaction levels.**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Dataset:**[**https://drive.google.com/drive/folders/1TGEc66YKbD443nslRi1bWgVd238gJCnb**](https://drive.google.com/drive/folders/1TGEc66YKbD443nslRi1bWgVd238gJCnb)

**The data is available in 8 csv files:**

1. **customers.csv**
2. **sellers.csv**
3. **order\_items.csv**
4. **geolocation.csv**
5. **payments.csv**
6. **reviews.csv**
7. **orders.csv**
8. **products.csv**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**The column description for these csv files is given below.**

**The customers.csv contain following features:**

|  |  |
| --- | --- |
| **Features** | **Description** |
| **customer\_id** | **ID of the consumer who made the purchase** |
| **customer\_unique\_id** | **Unique ID of the consumer** |
| **customer\_zip\_code\_prefix** | **Zip Code of consumer’s location** |
| **customer\_city** | **Name of the City from where order is made** |
| **customer\_state** | **State Code from where order is made (Eg. são paulo - SP)** |

**The sellers.csv contains following features:**

|  |  |
| --- | --- |
| **Features** | **Description** |
| **seller\_id** | **Unique ID of the seller registered** |
| **seller\_zip\_code\_prefix** | **Zip Code of the seller’s location** |
| **seller\_city** | **Name of the City of the seller** |
| **seller\_state** | **State Code (Eg. são paulo - SP)** |

**The order\_items.csv contain following features:**

|  |  |
| --- | --- |
| **Features** | **Description** |
| **order\_id** | **A Unique ID of order made by the consumers** |
| **order\_item\_id** | **A Unique ID given to each item ordered in the order** |
| **product\_id** | **A Unique ID given to each product available on the site** |
| **seller\_id** | **Unique ID of the seller registered in Target** |
| **shipping\_limit\_date** | **The date before which the ordered product must be shipped** |
| **price** | **Actual price of the products ordered** |
| **freight\_value** | **Price rate at which a product is delivered from one point to another** |

**The geolocations.csv contain following features:**

|  |  |
| --- | --- |
| **Features** | **Description** |
| **geolocation\_zip\_code\_prefix** | **First 5 digits of Zip Code** |
| **geolocation\_lat** | **Latitude** |
| **geolocation\_lng** | **Longitude** |
| **geolocation\_city** | **City** |
| **geolocation\_state** | **State** |

**The payments.csv contain following features:**

|  |  |
| --- | --- |
| **Features** | **Description** |
| **order\_id** | **A Unique ID of order made by the consumers** |
| **payment\_sequential** | **Sequences of the payments made in case of EMI** |
| **payment\_type** | **Mode of payment used (Eg. Credit Card)** |
| **payment\_installments** | **Number of installments in case of EMI purchase** |
| **payment\_value** | **Total amount paid for the purchase order** |

**The orders.csv contain following features:**

|  |  |
| --- | --- |
| **Features** | **Description** |
| **order\_id** | **A Unique ID of order made by the consumers** |
| **customer\_id** | **ID of the consumer who made the purchase** |
| **order\_status** | **Status of the order made i.e. delivered, shipped, etc.** |
| **order\_purchase\_timestamp** | **Timestamp of the purchase** |
| **order\_delivered\_carrier\_date** | **Delivery date at which carrier made the delivery** |
| **order\_delivered\_customer\_date** | **Date at which customer got the product** |
| **order\_estimated\_delivery\_date** | **Estimated delivery date of the products** |

**The reviews.csv contain following features:**

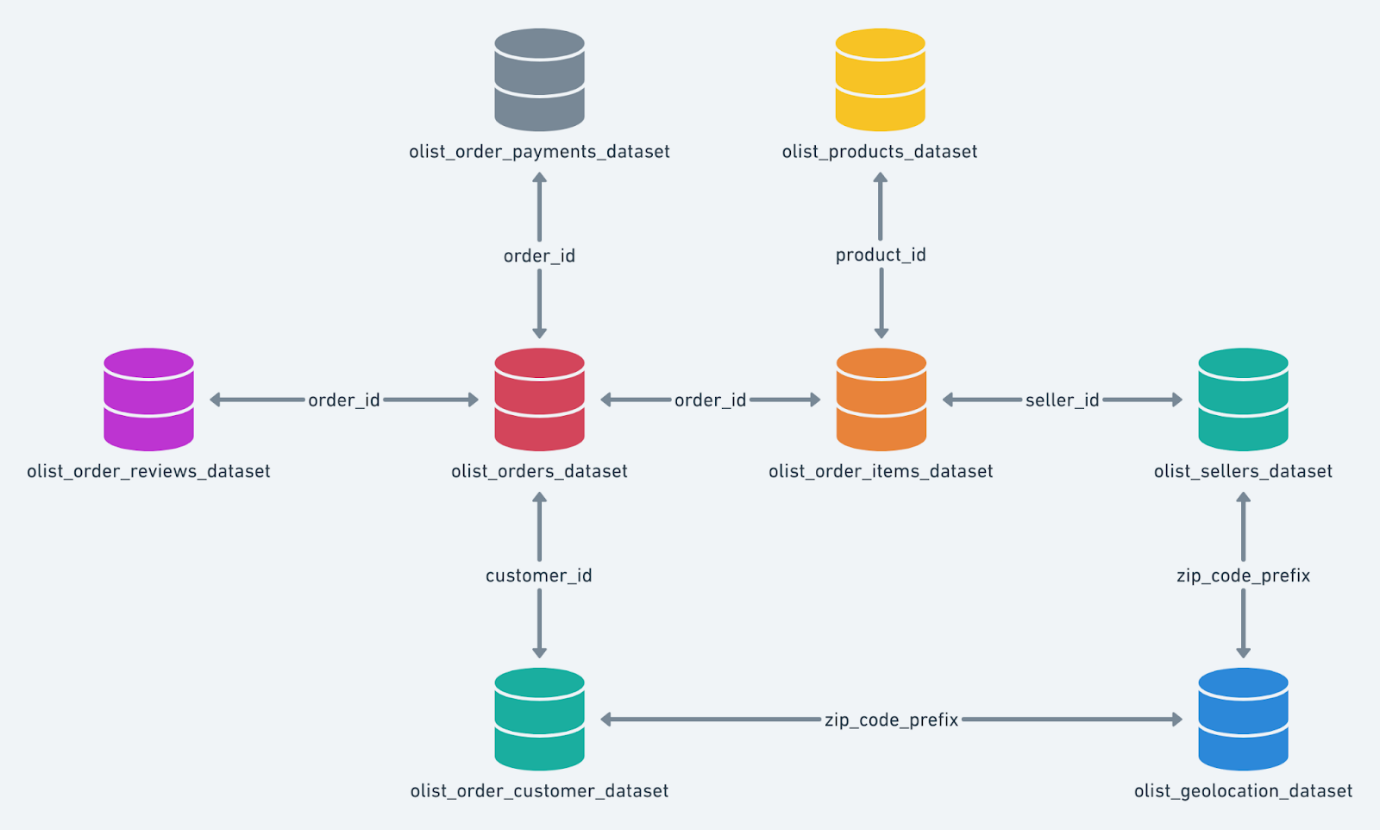
|  |  |
| --- | --- |
| **Features** | **Description** |
| **review\_id** | **ID of the review given on the product ordered by the order id** |
| **order\_id** | **A Unique ID of order made by the consumers** |
| **review\_score** | **Review score given by the customer for each order on a scale of 1-5** |
| **review\_comment\_title** | **Title of the review** |
| **review\_comment\_message** | **Review comments posted by the consumer for each order** |
| **review\_creation\_date** | **Timestamp of the review when it is created** |
| **review\_answer\_timestamp** | **Timestamp of the review answered** |

**The products.csv contain following features:**

|  |  |
| --- | --- |
| **Features** | **Description** |
| **product\_id** | **A Unique identifier for the proposed project.** |
| **product\_category\_name** | **Name of the product category** |
| **product\_name\_lenght** | **Length of the string which specifies the name given to the products ordered** |
| **product\_description\_lenght** | **Length of the description written for each product ordered on the site** |
| **product\_photos\_qty** | **Number of photos of each product ordered available on the shopping portal** |
| **product\_weight\_g** | **Weight of the products ordered in grams** |
| **product\_length\_cm** | **Length of the products ordered in centimeters** |
| **product\_height\_cm** | **Height of the products ordered in centimeters** |
| **product\_width\_cm** | **Width of the product ordered in centimeters** |

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Dataset schema:**

****

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

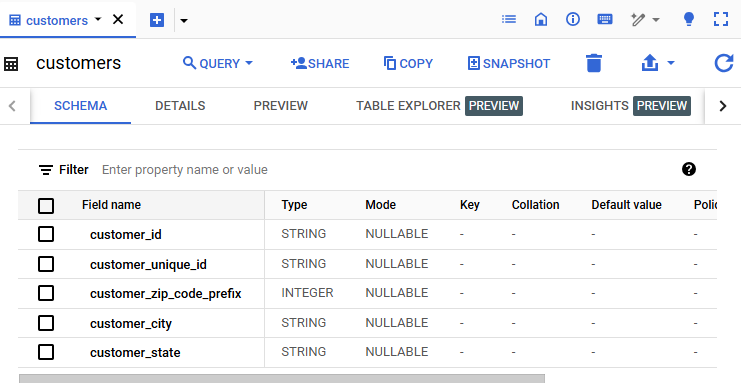
**Problem Statement:**

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

**Problem :** 1

Data type of all columns in the "customers" table.

**Output :**

****

**Problem : 2**

Get the time range between which the orders were placed.

**Query :**

select

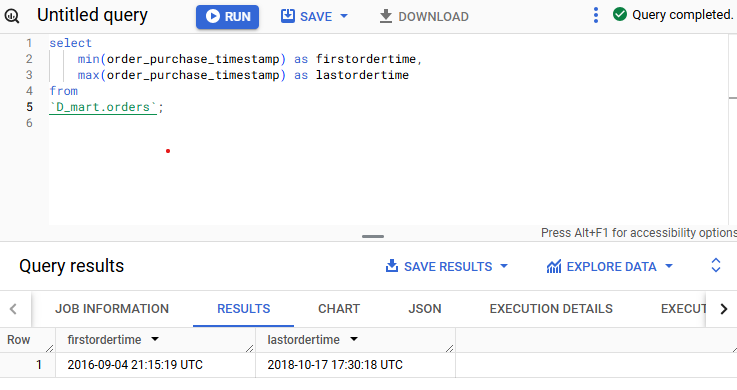
    min(order\_purchase\_timestamp) as firstordertime,

    max(order\_purchase\_timestamp) as lastordertime

from

`D\_mart.orders`;

**Output :**

****

**Problem : 3**

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

**Query :**

**Output :**

**II.In-depth Exploration:**

**Problem : 4**

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

**Query:**

select

  extract(month from order\_purchase\_timestamp) as order\_month,

  extract(year from order\_purchase\_timestamp) as order\_year,

  count(order\_id) as num\_orders

from

  D\_mart.orders

group by

  order\_year,

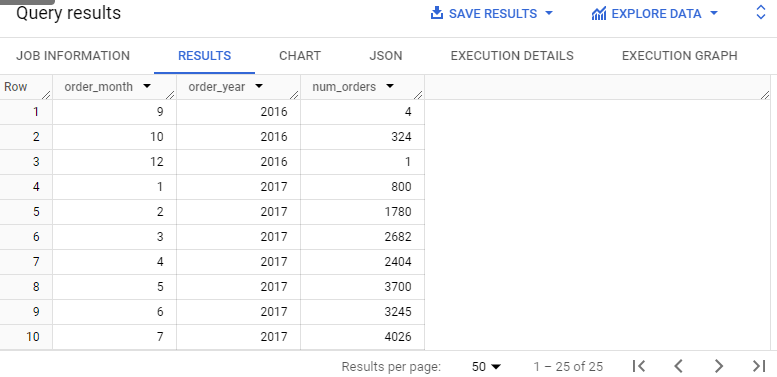
  order\_month

order by

  order\_year,

  order\_month;

**Output :**

****

**Problem : 5**

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

**Query :**

select

  extract(month from order\_purchase\_timestamp) as order\_month,

  avg(num\_orders) as average\_orders

from

  (

    select

      order\_purchase\_timestamp,

      count(order\_id) as num\_orders

    from

      D\_mart.orders

    group by

      order\_purchase\_timestamp

  ) as daily\_orders

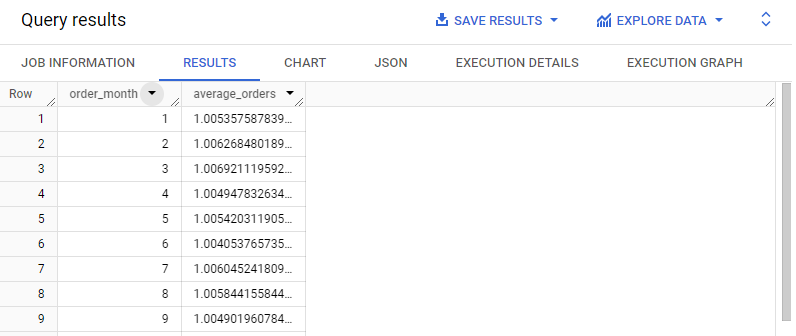
group by

  order\_month

order by

  order\_month;

**Output:**

****

**Problem : 6**

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 :**

**Output:**

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

**Problem : 7**

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

**Query :**

select

    customer\_state,

    extract(year from order\_purchase\_timestamp) as order\_year,

    extract(month from order\_purchase\_timestamp) as order\_month,

    count(order\_id) as num\_orders

from

    D\_mart.orders as o join D\_mart.customers as c on c.customer\_id = o.customer\_id

group by

    customer\_state,

    extract(year from order\_purchase\_timestamp),

    extract(month from order\_purchase\_timestamp)

order by

    customer\_state,

    order\_year,

    order\_month;

**Output :**

****

**Problem : 8**

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

**Query :**

select

    customer\_state,

    count(distinct customer\_id) as unique\_customers

from

    D\_mart.customers

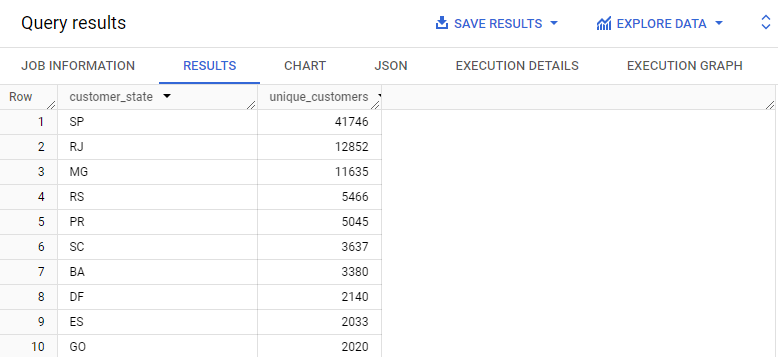
group by

    customer\_state

order by

    unique\_customers desc;

**Output :**

****

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

**Problem : 9**

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

**Query :**

with orders\_2017 as (

    select

        sum(payment\_value) as total\_2017

    from

        D\_mart.payments as p

    join

        D\_mart.orders as o on p.order\_id = o.order\_id

    where

        extract (year from o.order\_purchase\_timestamp) = 2017

        and extract (month from o.order\_purchase\_timestamp) between 1 and 8

),

orders\_2018 as (

    select

        sum(payment\_value) as total\_2018

    from

        D\_mart.payments as p

    join

        D\_mart.orders as o on p.order\_id = o.order\_id

    where

        extract (year from o.order\_purchase\_timestamp) = 2018

        and extract(month from o.order\_purchase\_timestamp) between 1 and 8

)

select

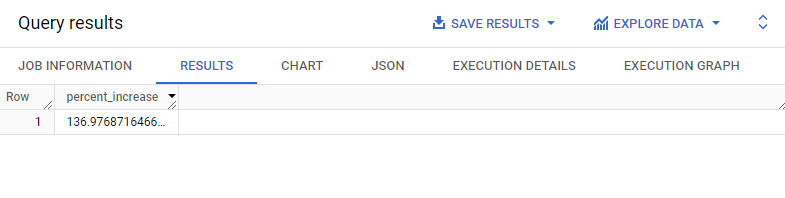
    ((o2018.total\_2018 - o2017.total\_2017) / o2017.total\_2017) \* 100 as percent\_increase

from

    orders\_2017 o2017,

    orders\_2018 o2018;

**Output :**

****

**Problem : 10**

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

**Query :**

select

    customer\_state,

    sum(payment\_value) as total\_order\_value,

    avg(payment\_value) as average\_order\_value

from D\_mart.customers as c join D\_mart.orders as o

on c.customer\_id = o.customer\_id

join D\_mart.payments as p on o.order\_id = p.order\_id

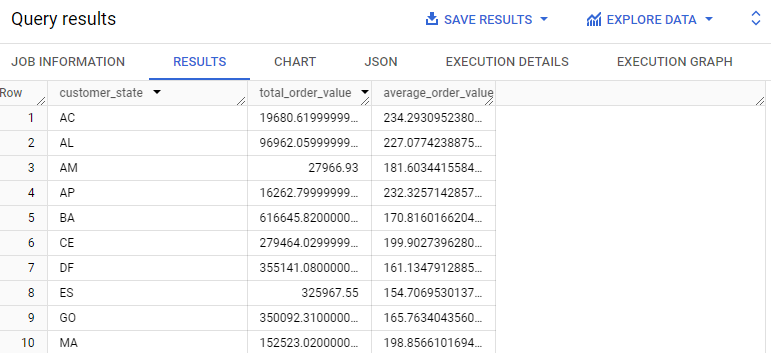
group by

    customer\_state

order by

    customer\_state;

**Output :**

****

**Problem : 11**

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

**Query :**

select

  customer\_state,

    sum(freight\_value) as total\_freight\_value,

    avg(freight\_value) as average\_freight\_value

from D\_mart.customers as c join D\_mart.orders as o

on c.customer\_id = o.customer\_id

join D\_mart.order\_items as oi on o.order\_id = oi.order\_id

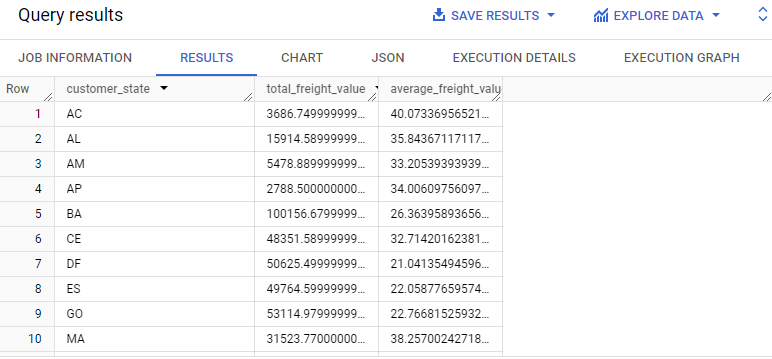
 group by

    customer\_state

order by

    customer\_state;

**Output :**

****

**V. Analysis based on sales, freight and delivery time.**

**Problem : 12**

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,

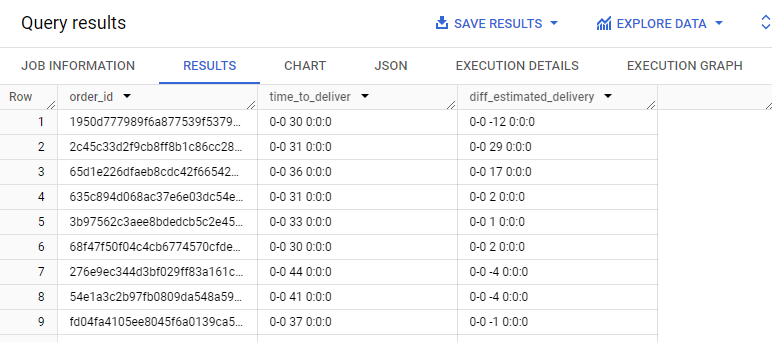
    date(order\_delivered\_customer\_date) - date(order\_purchase\_timestamp) as time\_to\_deliver,

    date(order\_estimated\_delivery\_date) - date(order\_delivered\_customer\_date) as

diff\_estimated\_delivery

from D\_mart.orders;

**Output:**

****

**Problem : 13**

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

**Query :**

**Output:**

**Problem : 14**

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

**Query :**

**Output :**

**Problem : 15**

Find out the top 5 states where the order delivery is really fast as compared to the estimated date of delivery. You can use the difference between the averages of actual & estimated delivery date to figure out how fast the delivery was for each state.

**Query :**

**Output :**

**VI. Analysis based on the payments:**

**Problem : 16**

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

**Query :**

select

    extract(year from order\_purchase\_timestamp) as year,

    extract(month from order\_purchase\_timestamp) as month,

    payment\_type,

    count(o.order\_id) as number\_of\_orders

from

    D\_mart.orders as o join

    D\_mart.payments as p on o.order\_id = p.order\_id

group by

    year,

    month,

    payment\_type

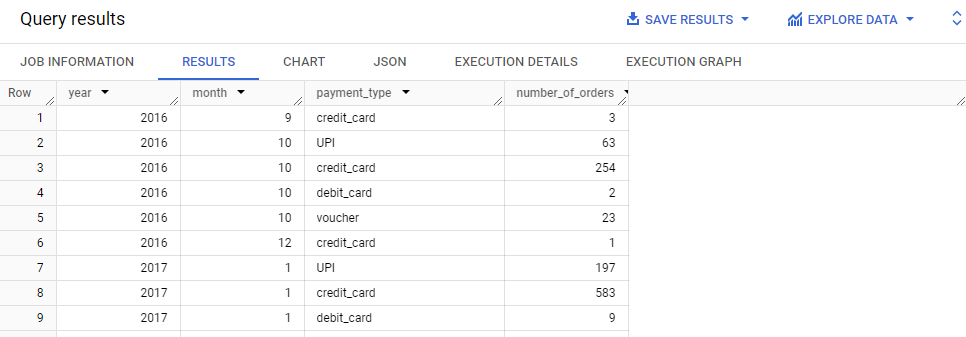
order by

    year,

    month,

    payment\_type;

**Output :**

****

**Problem : 17**

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

**Query :**

**Output :**