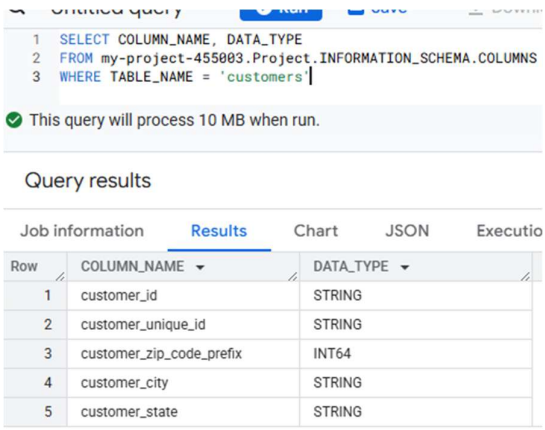
Q1) Import the dataset and do usual exploratory analysis steps like checking the structure & characteristics of the dataset:

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

```
SELECT COLUMN_NAME, DATA_TYPE
FROM my-project-455003.Project.INFORMATION_SCHEMA.COLUMNS
WHERE TABLE_NAME = 'customers'
```



```
1  SELECT COLUMN_NAME, DATA_TYPE
2  FROM my-project-455003.Project.INFORMATION_SCHEMA.COLUMNS
3  WHERE TABLE_NAME = 'customers'
```

This query will process 10 MB when run.

Query results

Job information | Results | Chart | JSON | Executio

Row	COLUMN_NAME	DATA_TYPE
1	customer_id	STRING
2	customer_unique_id	STRING
3	customer_zip_code_prefix	INT64
4	customer_city	STRING
5	customer_state	STRING

Recommendations & Insights

1. Almost all, all the columns are string data type except custome_zip_code_prefix , it indicates that table is dimension table.
2. Customer_id,customer_unique_id acts as a primary keys,for to access the data
3. custome_zip_code_prefix is int type but some zip codes might contains leading zeros,so for that it needs to be change
4. overall this table gives geographical information about the customers

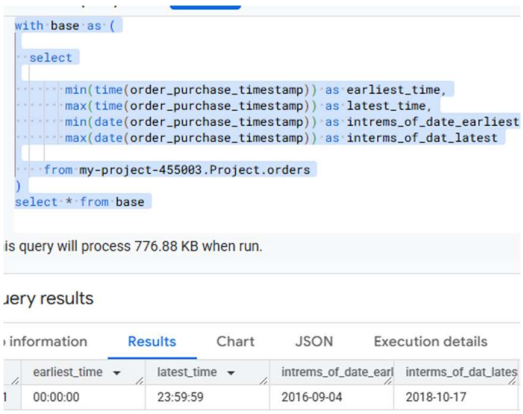
B) Get the time range between which the orders were placed.

```
with base as (

  select

      min(time(order_purchase_timestamp)) as earliest_time,
      max(time(order_purchase_timestamp)) as latest_time,
      min(date(order_purchase_timestamp)) as intrems_of_date_earliest,
      max(date(order_purchase_timestamp)) as interms_of_dat_latest

    from my-project-455003.Project.orders
)
select * from base
```



Insights

1. Orders can be placed at any time between 00:00:00 and 23:59:59, which means customers have 24-hour access to place orders.
2. The data spans from 2016-09-04 to 2018-10-17, covering over two years of order trends.
3. This allows for tracking seasonal trends, peak order times, and shifts in consumer purchasing habits.
4. Some industries may see higher orders in the morning vs. evening, while others might experience a surge in late-night transactions.
5. Knowing peak order times can help with staffing, marketing campaigns, and system optimizations.

Recommendations

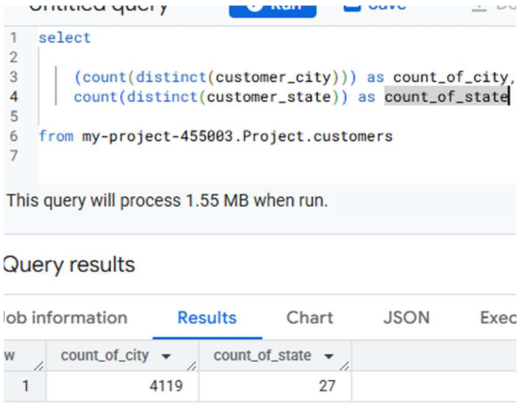
1. Look at order frequency by hour to identify peak times when customers are most active.
2. Use this information to schedule marketing promotions, customer support availability, and logistics operations efficiently.
3. Compare order placement patterns across different months/years to identify growth trends and seasonal peaks.
4. This can help forecast demand and plan stock availability.
5. If there are spikes in orders at specific hours, ensure that servers and payment gateways are optimized for high traffic.

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

```
select

    (count(distinct(customer_city))) as count_of_city,
    count(distinct(customer_state)) as count_of_state

from my-project-455003.Project.customers
```



Insights

1. 4,119 unique cities indicate a broad geographic distribution of customers, suggesting a well-established market presence.
2. This could mean strong national reach, diverse demand patterns, and opportunities for localized marketing strategies.
3. The fact that orders came from 27 states suggests that the business has nationwide appeal.
4. Some states may contribute higher order volumes, while others may have emerging markets with growth potential.
5. Some cities may have higher order density, while others have few but valuable customers.
6. The spread across states suggests differing logistics and delivery challenges depending on infrastructure and consumer concentration.

Recommendations

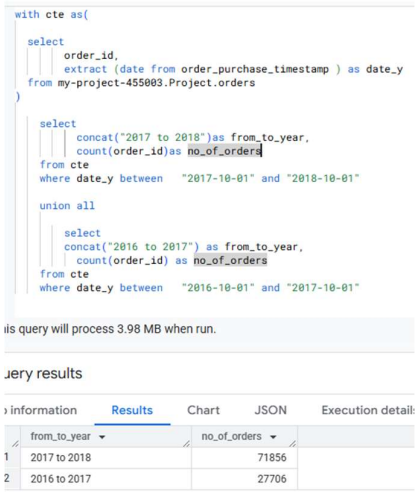
1. Segment cities based on order volume to prioritize marketing, inventory stocking, and faster delivery.
2. Focus more investment in states with the highest orders to drive growth.
3. With 4,119 cities in the network, optimizing delivery routes and regional warehouses can significantly reduce shipping costs and time
4. Identify cities with slower deliveries and improve supply chain efficiency
5. Personalized promotions by city/state can increase engagement and brand loyalty.
6. Target high-growth cities with exclusive deals and tailored product offerings.
7. Identify underserved states or regions and expand sales efforts in those areas.
8. Consider partnerships with local delivery providers to enhance accessibility and service speed.

Q2) In-depth Exploration:

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

```
with cte as(
  select
        order_id,
        extract (date from order_purchase_timestamp ) as date_y
```

```
  from my-project-455003.Project.orders
)
    select
          concat("2017 to 2018")as from_to_year,
          count(order_id)as no_of_orders
    from cte
    where date_y between   "2017-10-01" and "2018-10-01"
    union all
        select
        concat("2016 to 2017") as from_to_year,
          count(order_id) as no_of_orders
    from cte
    where date_y between   "2016-10-01" and "2017-10-01"
```



Insights

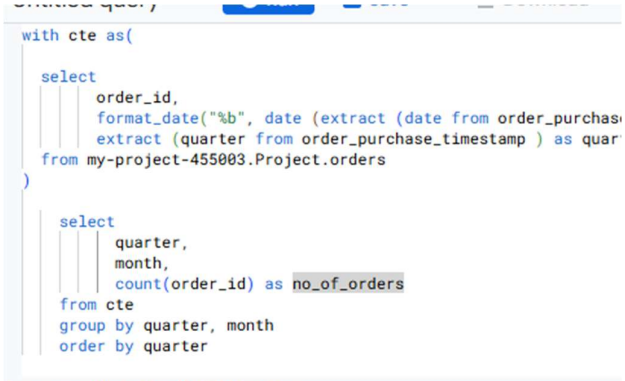
1. Orders increased from 27,706 in 2016-2017 to 71,856 in 2017-2018, indicating significant demand growth and business expansion.
2. The growth rate is approximately 159%, showing strong market adoption and increased customer engagement.

Recommendations

1. Analyse top-performing products, customer demographics, and demand patterns to further capitalize on growth.
2. Consider expanding into new regions or markets to sustain momentum.
3. Optimize warehouse distribution and delivery partners to ensure smooth order fulfillment.
4. Implement technology-driven inventory management for better stock forecasting.
5. Introduce loyalty programs and personalized promotions to encourage repeat purchases
6. Ensure fast, reliable support channels to keep customer satisfaction high.
7. Investigate whether the increase in orders is seasonal or consistent across months.
8. Plan peak-season strategies and adjust inventory levels accordingly.

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

```
with cte as(
  select
        order_id,
        format_date("%b", date (extract (date from order_purchase_timestamp
))) as month,
        extract (quarter from order_purchase_timestamp ) as quarter
  from my-project-455003.Project.orders
)
    select
          quarter,
          month,
          count(order_id) as no_of_orders
    from cte
    group by quarter, month
    order by quarter
```



uery results

information | Results | Chart | JSON | Execution detai

	quarter	month	no_of_orders
1	1	Feb	8508
2	1	Jan	8069
3	1	Mar	9893
4	2	Apr	9343
5	2	May	10573
6	2	Jun	9412
7	3	Sep	4305
8	3	Jul	10318
9	3	Aug	10843
0	4	Oct	4959
1	4	Nov	7544
2	4	Dec	5674

Insights

1. The highest order volume occurs in May (10,573), August (10,843), and July (10,318), indicating seasonal peaks in Q2 and Q3.
2. This suggests that customers are more active in purchasing during mid-year months—potentially aligning with sales, promotions, or consumer trends.
3. September (4,305), October (4,959), and December (5,674) have significantly lower order volumes, possibly due to seasonal slowdowns or post-holiday spending fatigue.
4. September has the lowest number of orders, potentially signaling a demand dip before holiday shopping kicks in.
5. Q2 (Apr-Jun) and Q3 (Jul-Sep) experience the highest order volumes, indicating strong purchasing trends during these periods.
6. Q4 (Oct-Dec) sees reduced demand, likely influenced by financial constraints post-holiday season or shifts in consumer behavior.

Recommendations

1. May, July, and August should be leveraged for heavy marketing efforts, discounts, and promotional campaigns.
2. Invest in inventory optimization to ensure stock availability during peak months.
3. Introduce off-season discounts and special offers in September and October to counteract slow sales periods.
4. Consider festive or end-of-year promotions for December to drive more orders during this dip.
5. Adjust stock levels to match seasonal demand, ensuring better resource allocation.
6. Streamline logistics and shipping strategies to handle increased orders during peak periods efficiently
7. Investigate whether holidays, industry events, or shopping trends impact seasonal spikes in orders.
8. If a correlation exists, adjust marketing calendars accordingly to capitalize on consumer trends.

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

```
      with cte as(
  select
        order_id,
        extract (hour from order_purchase_timestamp ) as hour
  from my-project-455003.Project.orders
),
base_1 as(
  select
          order_id,
          case
              when hour between  0 and 6
                      then "Dawn"
              when hour between 7 and 12
                      then "Mornings"
              when hour between 13 and 18
```

```
                    then "Afternoon"
               when hour between 19 and 23
                    then "Night"
          end as time_of_the_day
  from cte
)
    select
          time_of_the_day,
          count(order_id) as no_of_orders
    from base_1
    group by time_of_the_day
    order by count(order_id)
```

Insights

1. With 38,135 orders, the afternoon is the most active period for customers placing orders in Brazil.
2. This indicates that consumers likely make purchasing decisions after work breaks or during leisure hours.
3. 28,331 orders are placed at night, suggesting that many customers browse and buy products after their daily routine ends.
4. This trend aligns with evening shopping habits, where consumers have more time to explore deals.
5. 27,733 orders indicate a high volume of purchases in the morning, possibly from professionals placing orders before starting work.

6. Morning orders might be influenced by urgent buys or habitual shopping routines.
7. With only 5,242 orders, dawn shows the lowest engagement, which is expected since most customers are asleep or not actively shopping.

Recommendations

1. Run targeted ads and flash sales in the afternoon and night to capture high-order traffic.
2. Consider morning discounts for daily essentials and impulse purchases.
3. Send special offers during afternoons to capitalize on peak engagement.
4. Night-time reminders for abandoned carts could encourage more late-hour purchases.
5. Have support teams active during afternoons and evenings, when customers are most engaged.
6. Faster query resolution during peak shopping times may improve customer experience and retention.

Q3) Evolution of E-commerce orders in the Brazil region:

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

```
with cte as(
  select
        a.customer_state as state,
        format_date("%b", date (extract (date from
b.order_purchase_timestamp))) as month,
        b.order_id
  from my-project-455003.Project.customers a inner join my-project-
455003.Project.orders b on a.customer_id = b.customer_id

)
select
      state,
      month,
      count(order_id) as no_of_orders
from cte
group by state,month
order by
      case month
        WHEN 'Jan' THEN 1
        WHEN 'Feb' THEN 2
        WHEN 'Mar' THEN 3
        WHEN 'Apr' THEN 4
        WHEN 'May' THEN 5
        WHEN 'Jun' THEN 6
        WHEN 'Jul' THEN 7
        WHEN 'Aug' THEN 8
        WHEN 'Sep' THEN 9
        WHEN 'Oct' THEN 10
        WHEN 'Nov' THEN 11
        WHEN 'Dec' THEN 12
      end,no_of_orders desc
```

```sql
with cte as(

  select
      a.customer_state as state,
      format_date("%b", date (extract (date from b.order_purchase_timestamp))) as month,
      b.order_id
  from my-project-455003.Project.customers a inner join my-project-455003.Project.orders b on a.customer_id = b.customer_id

)
select
      state,
      month,
      count(order_id) as no_of_orders
from cte
group by state,month
order by
      case month
        WHEN 'Jan' THEN 1
        WHEN 'Feb' THEN 2
        WHEN 'Mar' THEN 3
        WHEN 'Apr' THEN 4
        WHEN 'May' THEN 5
        WHEN 'Jun' THEN 6
        WHEN 'Jul' THEN 7
        WHEN 'Aug' THEN 8
        WHEN 'Sep' THEN 9
        WHEN 'Oct' THEN 10
        WHEN 'Nov' THEN 11
        WHEN 'Dec' THEN 12
      end,no_of_orders desc
```

is query will process 10.81 MB when run.

uery results

information | Results | Chart | JSON | Execution details | Execution graph

	state	month	no_of_orders
1	SP	Jan	3351
2	RJ	Jan	990
3	MG	Jan	971
4	PR	Jan	443
5	RS	Jan	427
6	SC	Jan	345
7	BA	Jan	264
8	GO	Jan	164
9	ES	Jan	159
10	DF	Jan	151

Insights

1. RJ (Rio de Janeiro) and MG (Minas Gerais) consistently have high order numbers across multiple months, indicating strong demand and customer activity in these regions.
2. RS (Rio Grande do Sul) and PR (Paraná) also show steady order placements, particularly in August, May, and July.
3. November sees a spike in RJ (1,048 orders) and MG (943 orders), potentially influenced by holiday shopping or promotional events.
4. Orders dip in September and October, likely due to post-holiday spending slowdowns.
5. AC (Acre), AP (Amapá), RR (Roraima), TO (Tocantins), and AM (Amazonas) consistently show low order numbers across months, suggesting limited market penetration or logistical challenges in these regions.
6. AP and RR have the lowest order counts (8-11 per month), which might indicate low consumer engagement or geographical barriers affecting delivery efficiency
7. States with mid-range orders like RO (Rondônia) and SE (Sergipe) show slight seasonal variations, which could present opportunities for localized marketing strategies.
8. December sees increased activity across multiple states, suggesting end-of-year shopping trends influencing consumer behavior.

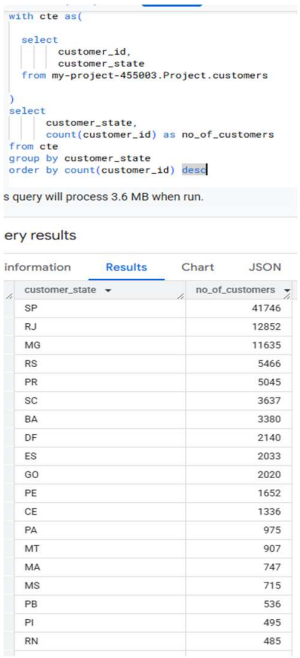
Recommendations

1. RJ, MG, RS, and PR should be the focus of marketing campaigns, seasonal discounts, and strategic inventory planning.

2. Boost promotional activity in November and December to align with peak shopping trends.
3. Strengthen delivery networks and promotional efforts in states like AC, AP, RR, TO, and AM to improve penetration.
4. Consider partnerships with local logistics providers for better service reach.
5. Ensure stock availability in peak-order months like November and August to meet rising demand.
6. Improve delivery efficiency in slower-order months to encourage consistent customer engagement.
7. Introduce mid-season promotions or bundled offers to keep sales momentum during traditionally slower months.
8. Analyze external trends (economic shifts, holiday patterns) influencing low order volumes in these months.

B) How are the customers distributed across all the states

```
with cte as(
  select
        customer_id,
        customer_state
  from my-project-455003.Project.customers
)
select
      customer_state,
      count(customer_id) as no_of_customers
from cte
group by customer_state
order by count(customer_id) desc
```



```
with cte as(
  select
        customer_id,
        customer_state
  from my-project-455003.Project.customers
)
select
        customer_state,
        count(customer_id) as no_of_customers
from cte
group by customer_state
order by count(customer_id) desc
```

s query will process 3.6 MB when run.

ery results

information | Results | Chart | JSON

customer_state	no_of_customers
SP	41746
RJ	12852
MG	11635
RS	5466
PR	5045
SC	3637
BA	3380
DF	2140
ES	2033
GO	2020
PE	1652
CE	1336
PA	975
MT	907
MA	747
MS	715
PB	536
PI	495
RN	485

Recommendations & Insights

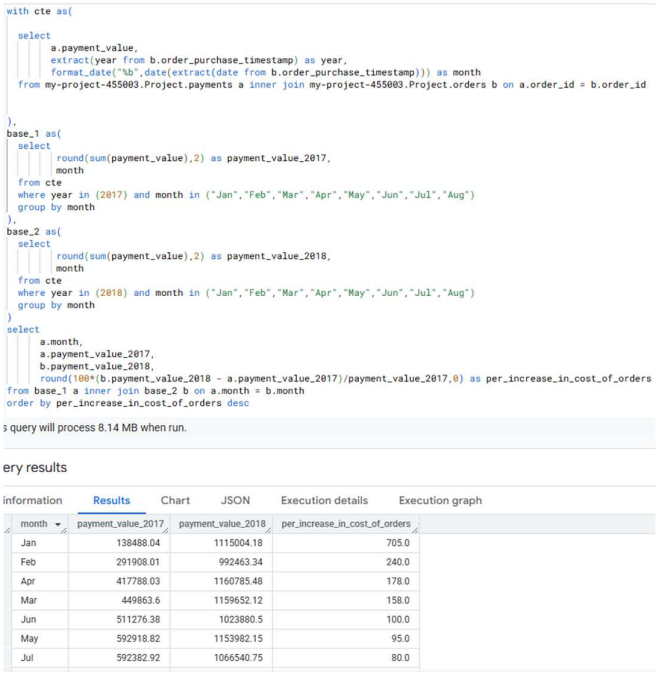
1. States like RR (Roraima), AP (Amapa), and AC (Acre) have fewer than 100 customers, suggesting very limited market presence there.
2. Among the bottom states, para (975 customers) and Mato Grosso (907) show relatively higher numbers, which could indicate growing market potential in otherwise under-penetrated regions.
3. Many of the lowest-ranking states in terms of customers are in the North and Northeast regions of Brazil – typically less urbanized and with lower population density.
4. Consider targeted campaigns or partnerships in underrepresented regions like Acre (AC), Roraima (RR), and Amapa (AP). These may represent untapped markets, especially if infrastructure and logistics allow.
5. Tailor messaging and promotions to regional needs and preferences. For example, use local dialects, festivals, and region-specific influencers in marketing.
6. States like MA, MS, and PB have decent customer bases that could be scaled up more easily than the very low-end states. Focused marketing and better retention efforts there may yield fast ROI.
7. In regions with less digital penetration, customer onboarding and trust-building through educational campaigns could help drive adoption.

Q4) 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).
You can use the "payment_value" column in the payments table to get the cost of orders.

```
with cte as(
  select
        a.payment_value,
        extract(year from b.order_purchase_timestamp) as year,
        format_date("%b",date(extract(date from
b.order_purchase_timestamp))) as month
  from my-project-455003.Project.payments a inner join my-project-
455003.Project.orders b on a.order_id = b.order_id
),
base_1 as(
  select
         round(sum(payment_value),2) as payment_value_2017,
         month
  from cte
  where year in (2017) and month in
("Jan","Feb","Mar","Apr","May","Jun","Jul","Aug")
  group by month
),
base_2 as(
  select
         round(sum(payment_value),2) as payment_value_2018,
         month
  from cte
```

```
  where year in (2018) and month in
("Jan","Feb","Mar","Apr","May","Jun","Jul","Aug")
  group by month
)
select
      a.month,
      a.payment_value_2017,
      b.payment_value_2018,
      round(100*(b.payment_value_2018 -
a.payment_value_2017)/payment_value_2017,0) as
per_increase_in_cost_of_orders
from base_1 a inner join base_2 b on a.month = b.month
order by per_increase_in_cost_of_orders desc
```



```
with cte as(

  select
      a.payment_value,
      extract(year from b.order_purchase_timestamp) as year,
      format_date("%b",date(extract(date from b.order_purchase_timestamp))) as month
  from my-project-455003.Project.payments a inner join my-project-455003.Project.orders b on a.order_id = b.order_id


),
base_1 as(
  select
      round(sum(payment_value),2) as payment_value_2017,
      month
  from cte
  where year in (2017) and month in ("Jan","Feb","Mar","Apr","May","Jun","Jul","Aug")
  group by month
),
base_2 as(
  select
      round(sum(payment_value),2) as payment_value_2018,
      month
  from cte
  where year in (2018) and month in ("Jan","Feb","Mar","Apr","May","Jun","Jul","Aug")
  group by month
)
select
      a.month,
      a.payment_value_2017,
      b.payment_value_2018,
      round(100*(b.payment_value_2018 - a.payment_value_2017)/payment_value_2017,0) as per_increase_in_cost_of_orders
from base_1 a inner join base_2 b on a.month = b.month
order by per_increase_in_cost_of_orders desc
```

s query will process 8.14 MB when run.

ery results

information | Results | Chart | JSON | Execution details | Execution graph

month	payment_value_2017	payment_value_2018	per_increase_in_cost_of_orders
Jan	138488.04	1115004.18	705.0
Feb	291908.01	992463.34	240.0
Apr	417788.03	1160785.48	178.0
Mar	449863.6	1159652.12	158.0
Jun	511276.38	1023880.5	100.0
May	592918.82	1153982.15	95.0
Jul	592382.92	1066540.75	80.0

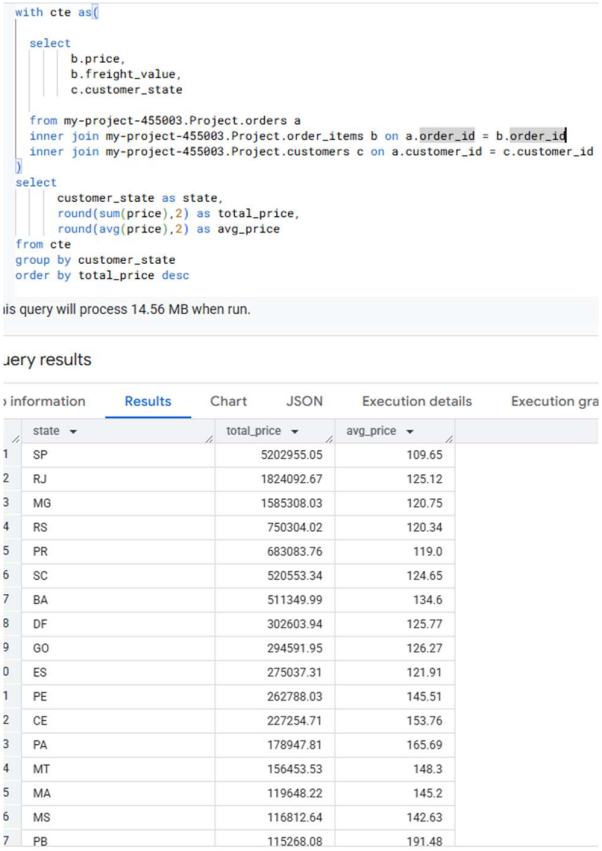
Recommendations & Insights

1. There is very significant growth from 2017 to 2018
2. There is dramatic increase in Jan and feb, indicating strong acceleration in customers spend in early of the year
3. May, jun , July are sugeesting steady demand
4. Since Jan and Feb saw explosive growth, invest in seasonal campaigns, discounts, or loyalty programs early in the year to continue that trend.
5. Compared to all the months , aug recorded lowest percentag (52%),it might be due to competition,customer retention issues,etc.,To understand this problem conduct customer satisfaction survey
6. May to July consistently see high order payments. Focus promotions, new product launches, or upselling strategies in these months to maximize returns.

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

```
with cte as(
  select
        b.price,
        b.freight_value,
        c.customer_state

  from my-project-455003.Project.orders a
  inner join my-project-455003.Project.order_items b on a.order_id =
b.order_id
  inner join my-project-455003.Project.customers c on a.customer_id =
c.customer_id
)
select
      customer_state as state,
      round(sum(price),2) as total_price,
      round(avg(price),2) as avg_price
from cte
group by customer_state
order by total_price desc
```



```
with cte as(

  select
        b.price,
        b.freight_value,
        c.customer_state

  from my-project-455003.Project.orders a
  inner join my-project-455003.Project.order_items b on a.order_id = b.order_id
  inner join my-project-455003.Project.customers c on a.customer_id = c.customer_id
)
select
      customer_state as state,
      round(sum(price),2) as total_price,
      round(avg(price),2) as avg_price
from cte
group by customer_state
order by total_price desc
```

is query will process 14.56 MB when run.

uery results

information | Results | Chart | JSON | Execution details | Execution gra

	state	total_price	avg_price
1	SP	5202955.05	109.65
2	RJ	1824092.67	125.12
3	MG	1585308.03	120.75
4	RS	750304.02	120.34
5	PR	683083.76	119.0
6	SC	520553.34	124.65
7	BA	511349.99	134.6
8	DF	302603.94	125.77
9	GO	294591.95	126.27
0	ES	275037.31	121.91
1	PE	262788.03	145.51
2	CE	227254.71	153.76
3	PA	178947.81	165.69
4	MT	156453.53	148.3
5	MA	119648.22	145.2
6	MS	116812.64	142.63
7	PB	115268.08	191.48

Insights

1. DF, GO, and ES lead in total price, suggesting high transaction volumes or higher demand in these regions.
2. PE and CE also have notable total prices, reinforcing their role in a strong trade network.

3. PB, AL, AC, and PA exhibit high average prices, indicating potential cost inefficiencies or premium pricing for goods/services.
4. PI, RO, and TO also show relatively high average prices, suggesting logistical or market challenges.
5. DF, GO, and ES maintain steady total price while keeping their average prices comparatively lower (~125–127), making them strong candidates for bulk operations or optimized cost structures.
6. States such as RR and AP have low total prices but relatively high average prices. This indicates smaller transaction volumes, possibly due to limited market size or high logistics costs.

Recommendations

1. States like PB, AL, AC, and PA may benefit from reducing overheads or negotiating better supplier and logistics terms.
2. Consider exploring alternative transportation methods or improving local sourcing to reduce shipping costs.
3. AP, RR, and AC may benefit from strategic marketing efforts to increase transaction volumes or diversify their offerings.
4. Exploring partnerships and increasing local demand for services/products could help improve revenue.
5. PA, RO, PI, and AL could focus on reducing supply chain inefficiencies or reassessing distribution models.
6. Improving infrastructure, using regional warehouses, or negotiating better transportation rates may lead to lower costs.

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

```
with cte as(

  select
        b.price,
        b.freight_value,
        c.customer_state

  from my-project-455003.Project.orders a
  inner join my-project-455003.Project.order_items b on a.order_id =
b.order_id
  inner join my-project-455003.Project.customers c on a.customer_id =
c.customer_id
)
select
      customer_state as state,
      round(sum(freight_value),2) as total_freight_price,
      round(avg(freight_value),2) as avg_freight_price
from cte
group by customer_state
order by total_freight_price desc
```

```sql
with cte as(

  select
      b.price,
      b.freight_value,
      c.customer_state

  from my-project-455003.Project.orders a
  inner join my-project-455003.Project.order_items b on a.order_id = b.order_id
  inner join my-project-455003.Project.customers c on a.customer_id = c.customer_id
)
select
      customer_state as state,
      round(sum(freight_value),2) as total_freight_price,
      round(avg(freight_value),2) as avg_freight_price
from cte
group by customer_state
order by total_freight_price desc
```

is query will process 14.56 MB when run.

uery results

information | Results | Chart | JSON | Execution details | Execution gra

	state	total_freight_price	avg_freight_price
1	SP	718723.07	15.15
2	RJ	305589.31	20.96
3	MG	270853.46	20.63
4	RS	135522.74	21.74
5	PR	117851.68	20.53
6	BA	100156.68	26.36
7	SC	89660.26	21.47
8	PE	59449.66	32.92
9	GO	53114.98	22.77
10	DF	50625.5	21.04
11	ES	49764.6	22.06
12	CE	48351.59	32.71
13	PA	38699.3	35.83
14	MA	31523.77	38.26
15	MT	29715.43	28.17
16	PB	25719.73	42.72

Insights

1. PE, GO, and DF have the highest total freight prices, indicating large freight volumes or high transportation costs.
2. PA, MA, and PB also show relatively high freight pricing.
3. PB, RR, RO, and AC have significantly higher average freight prices, suggesting expensive per-unit transportation costs.
4. RR, AP, AC, and AM exhibit lower total freight prices but high average freight prices, possibly due to difficult logistics (remote areas, specialized transport needs).
5. DF, GO, and MS show relatively lower average freight prices, indicating cost-effective transportation.

Recommendations

1. Focus on reducing costs in states with high average freight prices (PB, RR, RO, AC).
2. Improve supply chain efficiency in remote or hard-to-access areas
3. Areas with high total freight price but low average freight (GO, DF, MS) may benefit from bulk freight strategies and further negotiations with transport providers.
4. Explore alternative transport modes (rail, waterways) to reduce costs.
5. For states with extreme freight costs, consider local sourcing or distribution centers to minimize long-haul expenses

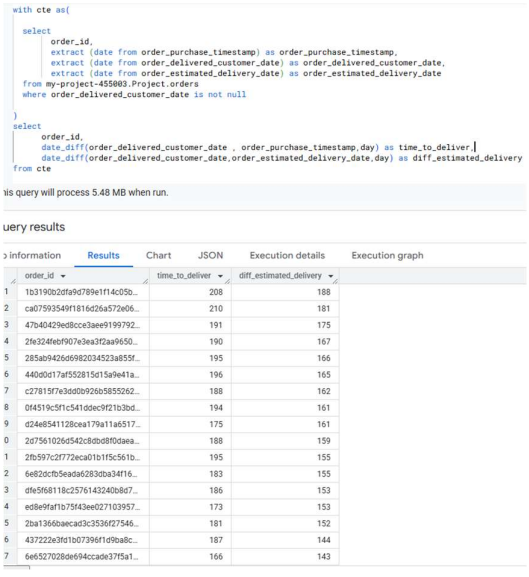
Q5) Analysis based on sales, freight and delivery time.

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.

```sql
with cte as(

  select
        order_id,
        extract (date from order_purchase_timestamp) as
order_purchase_timestamp,
        extract (date from order_delivered_customer_date) as
order_delivered_customer_date,
        extract (date from order_estimated_delivery_date) as
order_estimated_delivery_date
  from my-project-455003.Project.orders
  where order_delivered_customer_date is not null

)
select
      order_id,
      date_diff(order_delivered_customer_date , order_purchase_timestamp,day) as
time_to_deliver,
      date_diff(order_delivered_customer_date,order_estimated_delivery_date,day)
as diff_estimated_delivery
from cte
```



```sql
with cte as(
  select
      order_id,
      extract (date from order_purchase_timestamp) as order_purchase_timestamp,
      extract (date from order_delivered_customer_date) as order_delivered_customer_date,
      extract (date from order_estimated_delivery_date) as order_estimated_delivery_date
  from my-project-455003.Project.orders
  where order_delivered_customer_date is not null
)
select
      order_id,
      date_diff(order_delivered_customer_date , order_purchase_timestamp,day) as time_to_deliver,
      date_diff(order_delivered_customer_date,order_estimated_delivery_date,day) as diff_estimated_delivery
from cte
```

his query will process 5.48 MB when run.

uery results

information | Results | Chart | JSON | Execution details | Execution graph

	order_id	time_to_deliver	diff_estimated_delivery
1	1b3190b2dfa9d789e1f14c05b...	208	188
2	ca07593549f1816d26a572e06...	210	181
3	47b40429ed8cce3aee9199792...	191	175
4	2fe324febf907e3ea3f2aa9650...	190	167
5	285ab9426d6982034523a855f...	195	166
6	440d0d17af552815d15a9e41a...	196	165
7	c27815f7e3dd0b926b5855262...	188	162
8	0f4519c5f1c541ddec9f21b3bd...	194	161
9	d24e8541128cea179a11a6517...	175	161
0	2d7561026d542c8dbd8f0daea...	188	159
1	2fb597c2f772eca01b1f5c561b...	195	155
2	6e82dcfb5eada6283dba34f16...	183	155
3	dfe5f68118c2576143240b8d7...	186	153
4	ed8e9faf1b75f43ee027103957...	173	153
5	2ba1366baecad3c3536f27546...	181	152
6	437222e3fd1b07396f1d9ba8c...	187	144
7	6e6527028de694ccade37f5a1...	166	143

Insights

1. Some orders take more than 100 days to deliver, which is significantly high.
2. The longest delivery time recorded is 208 days, which suggests logistical inefficiencies or delays in certain regions
3. Many orders ,indicating they are arriving later than expected
4. The biggest delay observed is 188 days.
5. Some deliveries occur within 30-50 days, while others take over more than 100 days.
6. The variation could point to inconsistencies in supply chain management, distance factors, or vendor issues.

Recommendations

1. Identify common factors contributing to delays—whether it's shipping providers, inventory issues, or regional transport challenges.
2. Explore faster transport methods like air freight for urgent deliveries.
3. Adjust estimated delivery times to reflect realistic shipping durations, avoiding customer disappointment
4. Examine the root causes of extreme delays and prioritize solutions for these cases.
5. Consider regional warehouses to reduce transit times for distant orders.
6. Notify customers about potential delays in advance.
7. Offer compensation or expedited shipping to improve customer satisfaction.

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

```
with cte as(
  select
        b.freight_value,
        c.customer_state
  from my-project-455003.Project.orders a
  inner join my-project-455003.Project.order_items b on a.order_id =
b.order_id
  inner join my-project-455003.Project.customers c on a.customer_id =
c.customer_id
),base_1 as
(
select
      customer_state as state,
      round(avg(freight_value),2) as avg_freight_price
from cte
group by customer_state
order by avg_freight_price desc
limit 5),
base_2 as (
      select
      customer_state as state,
      round(avg(freight_value),2) as avg_freight_price
from cte
```

```
group by customer_state
order by avg_freight_price asc
limit 5
)
select * from base_1
union all
select * from base_2
```

```
with cte as(
  select
        b.freight_value,
        c.customer_state
  from my-project-455003.Project.orders a
  inner join my-project-455003.Project.order_items b on a.order_id = b.order_id
  inner join my-project-455003.Project.customers c on a.customer_id = c.customer_id
),base_1 as
(
select
      customer_state as state,
      round(avg(freight_value),2) as avg_freight_price
from cte
group by customer_state
order by avg_freight_price desc
limit 5),
base_2 as (
      select
      customer_state as state,
      round(avg(freight_value),2) as avg_freight_price
from cte
group by customer_state
order by avg_freight_price asc
limit 5
)
select * from base_1
union all
select * from base_2
```

is query will process 14.56 MB when run.

ıery results

information | Results | Chart | JSON | Execution details | Execution gra

state	avg_freight_price
RR	42.98
PB	42.72
RO	41.07
AC	40.07
PI	39.15
SP	15.15
PR	20.53
MG	20.63
RJ	20.96
DF	21.04

Insights

1. The top five states (RR, PB, RO, AC, PI) have significantly higher average freight prices, all exceeding 39 per unit.
2. This suggests higher transportation costs, which could be due to difficult terrain, long distances, or fewer freight carriers leading to higher prices.
3. The bottom five states (SP, PR, MG, RJ, DF) have the lowest average freight prices, with SP at only 15.15.
4. These states likely benefit from better infrastructure, shorter shipping distances, and competitive freight markets.
5. Remote states (RR, RO, AC) tend to have higher freight costs, possibly due to logistical challenges.
6. Industrial and highly developed regions (SP, RJ, DF) have lower costs, likely due to efficient transport networks

Recommendations

1. Negotiate better transport rates with logistics providers.
2. Explore alternative transport methods (rail, waterways) to minimize dependency on road transport.
3. Leverage cost savings in SP, PR, MG, RJ, and DF to build competitive pricing strategies.
4. Consider expanding operations in these states to maximize cost-effectiveness

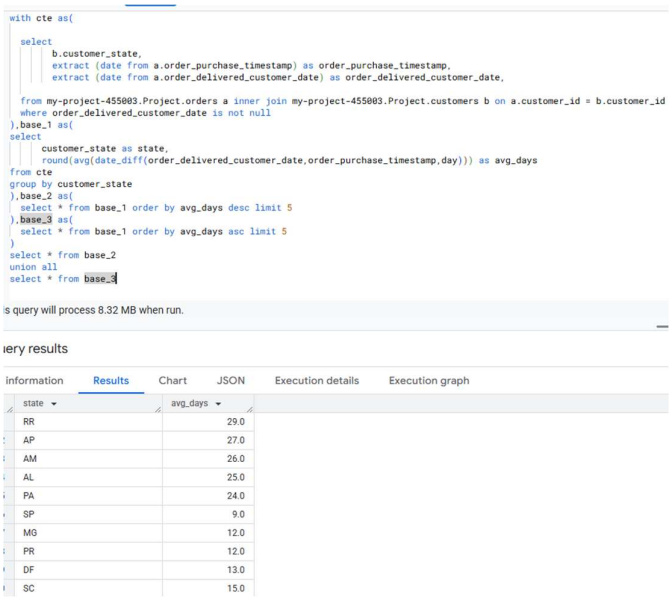
5. Governments or private players should invest in better roads, warehouses, and transport facilities in states like RR, RO, AC to reduce costs.
6. Reduce unnecessary delays and optimize truck loads for cost efficiency.

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

```
      with cte as(

  select
        b.customer_state,
        extract (date from a.order_purchase_timestamp) as
order_purchase_timestamp,
        extract (date from a.order_delivered_customer_date) as
order_delivered_customer_date,

  from my-project-455003.Project.orders a inner join my-project-
455003.Project.customers b on a.customer_id = b.customer_id
  where order_delivered_customer_date is not null
),base_1 as(
select
     customer_state as state,
     round(avg(date_diff(order_delivered_customer_date,order_purchase_times
tamp,day))) as avg_days
from cte
group by customer_state
),base_2 as(
  select * from base_1 order by avg_days desc limit 5
),base_3 as(
  select * from base_1 order by avg_days asc limit 5
)
select * from base_2
union all
select * from base_3
```



Insights

1. RR (29 days), AP (27 days), AM (26 days), AL (25 days), PA (24 days) have the highest average delivery days.
2. These delays may be due to challenging logistics, long distances, or inefficient transportation networks.
3. SP (9 days), MG (12 days), PR (12 days), DF (13 days), SC (15 days) show significantly shorter delivery times.
4. These regions likely benefit from better infrastructure, more efficient supply chains, and proximity to key transport hubs.
5. The gap between fastest (9 days) and slowest (29 days) is 20 days, indicating major inconsistencies in logistics efficiency across states
6. High delivery times in RR, AP, AM could lead to customer dissatisfaction, increased costs, and missed deadlines.

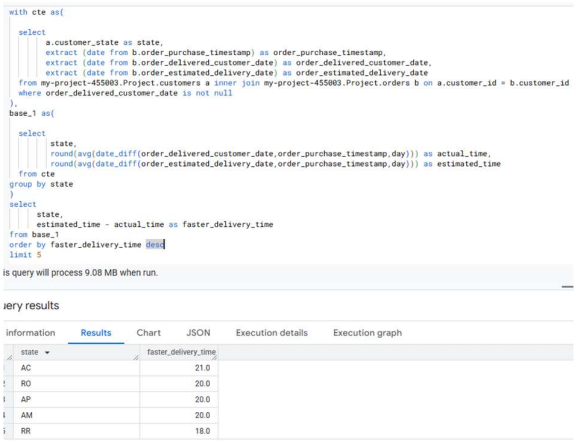
Recommendations

1. Setting up regional warehouses can shorten delivery times.
2. Work with logistics providers offering express shipping or optimized transport routes.
3. Identify whether delays are caused by traffic congestion, infrastructure issues, or transport bottlenecks.
4. SP, MG, PR, DF, SC should be prioritized for fast-moving consumer goods (FMCG) and urgent deliveries.
5. Refine estimated delivery dates based on historical trends to set realistic expectations.
6. Provide real-time tracking updates to customers to improve transparency.

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

```
with cte as(
  select
        a.customer_state as state,
        extract (date from b.order_purchase_timestamp) as
order_purchase_timestamp,
        extract (date from b.order_delivered_customer_date) as
order_delivered_customer_date,
        extract (date from b.order_estimated_delivery_date) as
order_estimated_delivery_date
  from my-project-455003.Project.customers a inner join my-project-
455003.Project.orders b on a.customer_id = b.customer_id
  where order_delivered_customer_date is not null
),
base_1 as(
  select
         state,
         round(avg(date_diff(order_delivered_customer_date,order_purchase
_timestamp,day))) as actual_time,
         round(avg(date_diff(order_estimated_delivery_date,order_purchase
_timestamp,day))) as estimated_time
  from cte
group by state
)
```

```sql
select
      state,
      estimated_time - actual_time as faster_delivery_time
from base_1
order by faster_delivery_time desc
limit 5
```



Insights

1. AC (21.0 days faster) leads in early deliveries. This suggests strong logistics efficiency or favorable infrastructure.
2. These states have nearly identical faster delivery times (**20 days**). This might indicate shared best practices or regional advantages.
3. RR (18.0 days faster) still performs well, though it's slightly behind the others. It could mean minor delays due to regional constraints.

Recommendations

1. Investigate AC to understand what's driving its success and replicate best practices in other regions.
2. Improve RR's performance by identifying bottlenecks causing the slight lag
3. Leverage insights from these states to implement predictive analytics across other regions.
4. Consider regional partnerships with local courier services to maintain efficiency.
5. Fast deliveries are great, but verify if customers are truly satisfied.
6. Consider collecting feedback on delivery accuracy, package conditions, and communication.

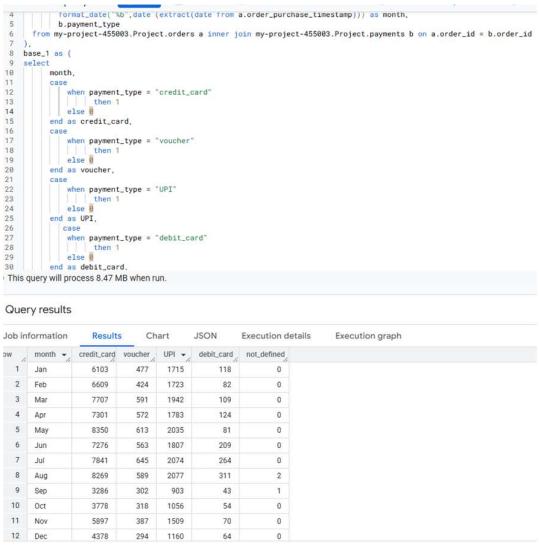
Q6) Analysis based on the payments:

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

```sql
with cte as (

  select
        format_date("%b",date (extract(date from
a.order_purchase_timestamp))) as month,
```

```
        b.payment_type
  from my-project-455003.Project.orders a inner join my-project-
455003.Project.payments b on a.order_id = b.order_id
),
base_1 as (
select
      month,
      case
          when payment_type = "credit_card"
                then 1
          else 0
      end as credit_card,
      case
          when payment_type = "voucher"
                then 1
          else 0
      end as voucher,
      case
          when payment_type = "UPI"
                then 1
          else 0
      end as UPI,
         case
          when payment_type = "debit_card"
                then 1
          else 0
      end as debit_card,
      case
          when payment_type = "not_defined"
                then 1
          else 0
      end as not_defined
  from cte
)
select
      month,
      sum(credit_card) as credit_card ,
      sum(voucher) as voucher,
      sum(UPI) as UPI,
      sum(debit_card) as debit_card,
      sum(not_defined) as not_defined
from base_1
group by month
order by
case month
        WHEN 'Jan' THEN 1
        WHEN 'Feb' THEN 2
        WHEN 'Mar' THEN 3
        WHEN 'Apr' THEN 4
        WHEN 'May' THEN 5
        WHEN 'Jun' THEN 6
        WHEN 'Jul' THEN 7
        WHEN 'Aug' THEN 8
        WHEN 'Sep' THEN 9
        WHEN 'Oct' THEN 10
        WHEN 'Nov' THEN 11
        WHEN 'Dec' THEN 12
      end
```



```
    format_date('%b',date (extract(date from a.order_purchase_timestamp))) as month,
    b.payment_type
from my-project-455003.Project.orders a inner join my-project-455003.Project.payments b on a.order_id = b.order_id
),
base_1 as (
select
    month,
    case
        when payment_type = "credit_card"
            then 1
        else 0
    end as credit_card,
    case
        when payment_type = "voucher"
            then 1
        else 0
    end as voucher,
    case
        when payment_type = "UPI"
            then 1
        else 0
    end as UPI,
    case
        when payment_type = "debit_card"
            then 1
        else 0
    end as debit_card,
```

This query will process 8.47 MB when run.

Query results

Job information | Results | Chart | JSON | Execution details | Execution graph

row	month	credit_card	voucher	UPI	debit_card	not_defined
1	Jan	6103	477	1715	118	0
2	Feb	6609	424	1723	82	0
3	Mar	7707	591	1942	109	0
4	Apr	7301	572	1783	124	0
5	May	8350	613	2035	81	0
6	Jun	7276	563	1807	209	0
7	Jul	7841	645	2074	264	0
8	Aug	8269	589	2077	311	2
9	Sep	3286	302	903	43	1
10	Oct	3778	318	1056	54	0
11	Nov	5897	387	1509	70	0
12	Dec	4378	294	1160	64	0

Insights

1. Credit card payments show a consistent upward trend, peaking in May (8,350) and August (8,269) before dropping significantly in September (3,286).
2. The lowest months for credit card usage are September and October, which may indicate seasonal factors or reduced spending activity.
3. UPI transactions steadily increase from 1,715 in January to 2,077 in August, indicating strong adoption of digital payments.
4. The peak UPI usage is in May (2,035) and August (2,077), reinforcing consumer preference for instant transactions.
5. Voucher payments fluctuate but remain between 294 (Dec) and 645 (Jul).
6. The highest redemption happens in July (645), possibly due to promotional campaigns or seasonal discounts.
7. Debit card transactions remain significantly lower than other payment methods, peaking at 311 in August, but never exceeding credit card or UPI transactions.
8. This indicates that users prefer credit cards or UPI over direct debit payments, possibly due to rewards or cashback offers.
9. A noticeable decline across all payment methods in September and October, with credit card transactions falling to nearly half their usual numbers.
10. This might indicate a seasonal dip in purchases—could be post-holiday exhaustion or reduced consumer spending in these months.

Recommendations

1. Focus on cashback or EMI options to attract more users.

2. Promote offers specifically in peak months like May and August when transactions are the highest.
3. Given UPI's popularity, partner with digital wallets or banks to provide discounts for UPI payments.
4. Reduce reliance on manual payment processing by integrating quick UPI checkout options.
5. Identify why transactions drop in these months and introduce seasonal campaigns or promotions.

6. Offer festive or pre-holiday discounts to encourage spending.
7. Increase voucher promotions in months like June and July, where redemption rates are high.
8. Ensure easy redeemability across multiple platforms to maintain consumer interest.

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

```
with cte as (

  select
        order_id,
        payment_installments
  from my-project-455003.Project.payments
  where payment_installments <> 0
)
select
      payment_installments,
      count(order_id) as no_of_orders
from cte
group by payment_installments
order by payment_installments
```

```sql
with cte as (

  select
      order_id,
      payment_installments
  from my-project-455003.Project.payments
  where payment_installments <> 0
)
select
      payment_installments,
      count(order_id) as no_of_orders
from cte
group by payment_installments
order by payment_installments
```

This query will process 4.16 MB when run.

Query results

Job information | Results | Chart | JSON

w	payment_installment	no_of_orders
1	1	52546
2	2	12413
3	3	10461
4	4	7098
5	5	5239
6	6	3920
7	7	1626
8	8	4268
9	9	644
10	10	5328
11	11	23
12	12	133
13	13	16
14	14	15
15	15	74
16	16	5
17	17	8
18	18	27
19	20	17

Insights

1. 52,546 orders (highest count) are paid in full (1 installment), showing a clear preference for one-time payments over installment-based purchases.
2. Consumers might be avoiding long-term commitments or benefiting from discounts on upfront payments.
3. The number of orders drops significantly beyond 3 installments, indicating fewer people prefer extended payment plans.

4. Installments beyond 6 have much lower adoption, suggesting limited interest in long-term financing options.
5. There are still 1,626+ orders using 7-10 installments, indicating demand for financing options.
6. Higher installment plans (12-24 months) have fewer users, likely due to higher interest rates or long repayment commitments.

Recommendations

1. Given the strong preference for single payments, businesses can offer discounts for full payments to encourage quicker conversions.

2. Since multi-installment orders drop sharply after 3 installments, retailers can introduce 0% interest EMI plans for purchases needing financing.
3. Since installments up to 6 months still show demand, companies can promote short-term EMI options to drive conversions without long-term risk.
4. Low adoption of 12+ installment plans suggests potential customers might be unaware or hesitant.
5. Marketing campaigns educating users about benefits like low-interest EMIs or flexible payments could increase adoption.