TARGET_BUSINESS CASE OF SQL

What does 'good' look like?

1. **Import 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.

   Query:

```
select column_name,

    data_type
    from `scaler-dsml-sql-452804.target.INFORMATION_SCHEMA.COLUMNS`
    WHERE table_name='customers'
```

output:

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

INSIGHTS:

1. Multiple customer_ids can exist for the same customer_unique_id

2. customer_zip_code_prefix (INT64) :

Indicates the geographical region (first digits of the zip code).Useful for regional sales analysis, delivery logistics, and demand forecasting.

Recommendations:

1. can Analyze customer_zip_code_prefix and customer_city to find high-density customer areas.

2.can Generate dashboards to compare sales across different cities.

2. Get the time range between which the orders were placed.
   Query:

```
select min(order_purchase_timestamp) as first_order_date,
       max(order_purchase_timestamp) as last_order_date
         from `target.orders`
```

Output:

Row	first_order_date	last_order_date
1	2016-09-04 21:15:19 UTC	2018-10-17 17:30:18 UTC

Insights:

1.by observing data we get to know that data spans of 2 years.
2. A timeline can show how quickly the platform scaled — e.g., low orders in 2016 vs. rapid rise in 2017.

Recommendations:

1. we can use this to analyze **year-over-year** growth, peak seasons, and customer retention trends.

2. we can Align promotions with periods of historical high demand like summer sales.

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

Query:

```
select count(distinct c.customer_city) as city_count,
       count (distinct c.customer_state) as state_count
       from `target.customers` c join `target.orders` o on c.customer_id=o.customer_id
```

Output:

Row	city_count	state_count
1	4119	27

Insights:

1. By looking into output,we can say orders are placed from 4119 different cities across 27 states.

Recommendation:

1. Focus on strengthening delivery times and return handling in high-density cities related to logistics.

In-depth Exploration:

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

```
WITH yearly_orders AS (
  SELECT
    EXTRACT(YEAR FROM order_purchase_timestamp) AS year,
    COUNT(*) AS total_orders,

  FROM `target.orders`
  GROUP BY year
)

select year,total_orders,
       lag(total_orders,1) over(order by year) as prev_year_order_count,
       (total_orders-lag(total_orders,1) over(order by year)) as difference_in_orders
       from yearly_orders

       order by year
```

Output:

Row	year	total_orders	prev_year_order_cou	difference_in_orders
1	2016	329	null	null
2	2017	45101	329	44772
3	2018	54011	45101	8910

Insights:

1. From 329 orders in 2016 to 45,101 orders in 2017 indicates 2016 was likely a pilot or test launch year and 2017 was the real operational launch**,** and the platform took off rapidly.
2. 2018 continued to grow with 8,910 more orders**,** Not as explosive as 2017, but still positive and healthy**.**
3. Shows market adoption is stabilizing**,** business is entering a mature scaling phase**.**

Recommendations:

1. 2018's smaller growth suggests potential market saturation or rising competition.

   For that, Need to introduce subscription models, bundle offers, or exclusive

2.products.The sharp growth in 2017 points to the success of strategies like marketing, UX improvements, and seller onboarding.
Reapplying these approaches in 2018 could help sustain growth and achieve similar order volume increases.

---

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

```
with olc as
(select extract(year from order_purchase_timestamp) as year,
        extract(month from order_purchase_timestamp) as month_number,
        FORMAT_TIMESTAMP('%B', order_purchase_timestamp) AS month_name,
       CASE
        WHEN EXTRACT(MONTH FROM order_purchase_timestamp) IN (12,1,2) THEN 'Winter'
        WHEN EXTRACT(MONTH FROM order_purchase_timestamp) IN (3,4,5) THEN 'Spring'
        WHEN EXTRACT(MONTH FROM order_purchase_timestamp) IN (6,7,8) THEN 'Summer'
```

```
ELSE 'Fall'
END AS season,
count(*) as total_orders
from `target.orders`
group by 1,2,3,4
)
select year,month_number,month_name,season,total_orders,
       sum(total_orders) over(partition by season,year) as sesonality_orders_count
       from olc
       order by 1,2
```

Output:

Row	year	month_number	month_name	season	total_orders	sesonality_orders_count
1	2016	9	September	Fall	4	328
2	2016	10	October	Fall	324	328
3	2016	12	December	Winter	1	1
4	2017	1	January	Winter	800	8253
5	2017	2	February	Winter	1780	8253
6	2017	3	March	Spring	2682	8786
7	2017	4	April	Spring	2404	8786
8	2017	5	May	Spring	3700	8786
9	2017	6	June	Summer	3245	11602
10	2017	7	July	Summer	4026	11602

Insights:

1. Winter seasons consistently show high order volumes, This indicates strong customer activity during the holiday and New Year period.
2. Spring seasons maintain steady and reliable order volumes, reflecting consistent user engagement and possibly a stable customer base.
3. Fall seasons tend to have lower performance compared to 2017 year, indicating a potential drop in customer activity or engagement during this time.

Recommendations:

1. Plan large-scale marketing campaigns and product launches during summer to maximize sales, as it consistently shows the highest engagement.
2. Utilize the winter season for festive promotions, limited-time offers, and re-engagement strategies, capitalizing on naturally high customer interest.
3. Focus on improving fall performance through early-bird holiday campaigns, exclusive discounts.

---

3. 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

Ans:
Query:

```
select
   case
   when extract(hour from order_purchase_timestamp) in (0,1,2,3,4,5,6) then 'Dawn'
   when extract(hour from order_purchase_timestamp) in (7,8,9,10,11,12) then 'Morning'
   when extract(hour from order_purchase_timestamp) in (13,14,15,16,17,18) then 'Afternoon'
      else 'night' end as day_time,
   count(*) as order_placed
   from `target.orders`
   group by day_time
   order by order_placed desc
```

Output:

Row	day_time	order_placed
1	Afternoon	38135
2	night	28331
3	Morning	27733
4	Dawn	5242

Insights:

1. Afternoon has the highest order placement rate
2. Less orders placed in dawn time due to sleeping time.

Recommendations:

1.better to maintain good servers in afternoon times.

2 Optimize push notifications or delivery offers in morning,night times for attention

---

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

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

Answer:

Query:

```
select
   extract(year from o.order_purchase_timestamp) as year,
   extract(month from o.order_purchase_timestamp) as month,
   c.customer_state,
   count(o.order_purchase_timestamp) as total_orders_count
   from `target.orders` o join `target.customers` c on c.customer_id=o.customer_id
   group by 1,2,3
   order by 1,2
```

Output:

Row	year	month	customer_state	total_orders_count
1	2016	9	RR	1
2	2016	9	RS	1
3	2016	9	SP	2
4	2016	10	AL	2
5	2016	10	BA	4
6	2016	10	CE	8
7	2016	10	DF	6
8	2016	10	ES	4
9	2016	10	GO	9
10	2016	10	MA	4

Results per page: 50 1 – 50 of 565

Insights:

1. Certain states like São Paulo, Rio de Janeiro, and Minas Gerais consistently show high order volumes month over month, indicating strong and stable customer bases.

2. Some states show sharp rises or drops in specific months, which could be due to seasonal trends, regional promotions, or supply chain factors.

3. Smaller or less active states may have stagnant or fluctuating trends, suggesting untapped potential or logistical challenges.

Recommendations:

1.Focus marketing efforts and promotions in high-performing states to strengthen loyalty and boost repeat orders.

2.Investigate the causes behind spikes or drops in monthly orders in specific states to identify successful campaigns or underlying issues.

3.Use this MoM data to plan inventory and logistics operations more efficiently across states by forecasting demand patterns.

---

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

Answer:

Query:

```
select count( distinct customer_id) as total_customers,
       customer_state
       from `target.customers`
       group by 2
       order by 1 desc
```

Output:

Row	total_customers	customer_state
1	41746	SP
2	12852	RJ
3	11635	MG
4	5466	RS
5	5045	PR
6	3637	SC
7	3380	BA
8	2140	DF
9	2033	ES
10	2020	GO

Results per page: 50 1 – 27 of 27

Insights:

1. The majority of customers are concentrated in a few key states like São Paulo (SP), Rio de Janeiro (RJ), and Minas Gerais (MG), which are also Brazil's most urbanized and economically active regions.

2. Northern and some central-western states show lower customer counts, indicating either lower digital penetration or logistical challenges in these areas.

Recommendations:

1. Prioritize delivery optimization and customer support in high-density states to maintain satisfaction and boost lifetime value.

2. Consider targeted digital marketing campaigns in low-penetration states to increase customer base and brand presence.

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

1. 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.

   Answer:

   Query:

```
    with Payment_year as (
                select extract(year from o.order_purchase_timestamp) as year,
                       sum(p.payment_value) as payment_val
                       from `target.orders` o
                       join `target.payments` p on o.order_id = p.order_id
                  where extract(month from o.order_purchase_timestamp) in (1,2,3,4,5,6,7,8)
                  group by year
)
select max(case when year = 2017 then payment_val end) as payment_2017_year,
       max(case when year = 2018 then payment_val end) as payment_2018_year,
       ROUND(
           100 * (MAX(CASE WHEN year = 2018 THEN payment_val END) -
                  MAX(CASE WHEN year = 2017 THEN payment_val END)) /
                  MAX(CASE WHEN year = 2017 THEN payment_val END), 2
       ) as percent_increase
from Payment_year
```

Output:

Row	payment_2017_year	payment_2018_year	percent_increase
1	3669022.119999...	8694733.840000...	136.98

Insights:

1. From January to August, there was a significant increase in the total payment value made by customers in 2018 compared to 2017.
2. This implies better customer engagement, higher-value purchases, or possibly an increase in the number of successful transactions.

Recommendations:

1. Focus on analyzing the strategies implemented during early 2018, such as improved seller onboarding, promotions, or delivery efficiency.

2. Replicating these approaches in future cycles could further boost payment values. Additionally, exploring product categories or customer segments that contributed the most to the payment spike may help identify high-value opportunities.

---

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

Answer:

Query1:

```
SELECT
    c.customer_state,
    round(SUM(p.payment_value), 2) AS total_price,
    round(AVG(p.payment_value), 2) AS avg_price
FROM
    `target.payments` p
JOIN
    `target.orders` o ON p.order_id = o.order_id
JOIN
    `target.customers` c ON o.customer_id = c.customer_id
GROUP BY
    c.customer_state
ORDER BY
    c.customer_state;
```

Query2:

```
 WITH StateTotals AS (
    SELECT
        c.customer_state,
        SUM(p.payment_value) OVER (PARTITION BY c.customer_state) AS total_price,
        AVG(p.payment_value) OVER (PARTITION BY c.customer_state) AS avg_price
    FROM
        `target.payments` p
    JOIN
        `target.orders` o ON p.order_id = o.order_id
    JOIN
        `target.customers` c ON o.customer_id = c.customer_id
)
SELECT DISTINCT
    customer_state,
    total_price,
    avg_price
FROM
    StateTotals
ORDER BY
    customer_state;
```

Output:

Row	customer_state	total_price	avg_price
1	AC	19680.62	234.2930952380...
2	AL	96962.06	227.0774238875...
3	AM	27966.93	181.6034415584...
4	AP	16262.8	232.3257142857...
5	BA	616645.82	170.8160166204...
6	CE	279464.03	199.9027396280...
7	DF	355141.08	161.1347912885...
8	ES	325967.55	154.7069530137...
9	GO	350092.31	165.7634043560...
10	MA	152523.02	198.8566101694...

Results per page: 50 1 – 27 of 27

Insights:

1. State Variability in Total Order Value:

   Total order value may differ sharply among states, suggesting that certain states have a greater contribution to total revenue than others. States with higher total values might have population density, income, or a more active customer base.

2. Discrepancy in Average Order Price:

   If the average order value varies significantly from one state to another, it indicates that specific states can have disparate purchase patterns or product tastes. States with lower averages can be more price-conscious, and states with higher averages can be purchasing more high-end products.

3. Outliers in High Total Value Orders:

   There could be some outliers in highly total order value states because of a few high-value, high-ticket orders. This might distort the average price and total values and make some states look disproportionately large in revenue terms.

Recommendations:

1. Target High Revenue States with Focused Campaigns:

   For high total order states, spend on targeted advertising campaigns to keep and expand such customer bases. Providing loyalty schemes, special promotions, or exclusive offers might retain and even enhance business in such areas.

2. Make Pricing Strategies Optimized for Low Average Order States:

   In low average order price states, you may consider bundling products or developing tiered pricing strategies to boost the average order value. Providing discounts on bulk purchases or cross-selling complementary products may encourage customers to spend more.

3. Examine and Resolve Order Discrepancies:

   Examine if outliers (those states with unusually high total values) are caused by a few large orders or customer anomalies. If so, modifying business strategies for long-term, consistent revenue generation in such states could be used to even out these imbalances.

---

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

Answer:

Query 1:

```
SELECT

        c.customer_state,
        round(SUM(oi.freight_value),2) AS total_price,
        round(AVG(oi.freight_value),2) AS avg_price
        FROM  `target.order_items` oi
               JOIN `target.orders` o ON oi.order_id = o.order_id
               JOIN `target.customers` c ON o.customer_id = c.customer_id
        GROUP BY c.customer_state
        ORDER BY c.customer_state;
```

Query 2:

```
WITH StateTotals AS (
    SELECT
        c.customer_state,
        SUM(oi.freight_value) OVER (PARTITION BY c.customer_state) AS total_price,
        AVG(oi.freight_value) OVER (PARTITION BY c.customer_state) AS avg_price
    FROM
        `target.order_items` oi
    JOIN
        `target.orders` o ON oi.order_id = o.order_id
    JOIN
        `target.customers` c ON o.customer_id = c.customer_id
)
SELECT DISTINCT
    customer_state,
    total_price,
    avg_price
FROM
    StateTotals
ORDER BY
    customer_state;
```

Output:

Row	customer_state	total_price	avg_price
1	AC	3686.75	40.07
2	AL	15914.59	35.84
3	AM	5478.89	33.21
4	AP	2788.5	34.01
5	BA	100156.68	26.36
6	CE	48351.59	32.71
7	DF	50625.5	21.04
8	ES	49764.6	22.06
9	GO	53114.98	22.77
10	MA	31523.77	38.26

Results per page: 50 1 – 27 of 27

Insights:

1. SP Dominates in Total Freight (₹718K)
   Sao Paulo (SP) is well ahead of others when it comes to total freight value, suggesting heavy order volume and a probable central location for logistics.

2. RR, PB, and RO Show High Average Freight per Order
   States such as RR (₹42.98), PB (₹42.72), and RO (₹41.07) have relatively much higher freight cost per order, indicating longer delivery distance or challenging logistics for these states.

3. Low Freight in States Such as SP, DF, and MG
   States such as SP (₹15.15), DF (₹21.04), and MG (₹20.63) have low average freight, indicating improved delivery efficiency, perhaps due to improved logistics infrastructure or closeness to suppliers.

Recommendations:

1. Target Campaigns to High-Volume, Low-Value States like SP, MG, RJ are enormous in revenues but low in average spends.
   Upselling or bundling strategies can boost avg. order size.

2. Create Low-Revenue, High-AOV Areas RR, PB, RO indicate that customers are spending more per transaction.

5. 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.
   Do this in a single query.

   You can calculate the delivery time and the difference between the estimated & actual delivery date using the given formula:
   - **time_to_deliver** = order_delivered_customer_date - order_purchase_timestamp
   - **diff_estimated_delivery** = order_delivered_customer_date - order_estimated_delivery_date

Answer:

Query:

```
    select 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 `target.orders`
```

Output:

Row	time_to_deliver	diff_estimated_delivery
1	35	-16
2	30	-28
3	30	12
4	54	36
5	43	-6
6	36	-14
7	29	-20
8	30	-19
9	44	-5
10	68	18

Results per page: 50 1 – 50 of 99441

Insights:

1. Wide Range in Delivery Times can be seen in time_to_deliver column shows wide distribution. Orders are delivered as early as 1-3 days, and as late as more than 20 days.

2. Frequent Misses in Estimated Delivery like Most diff_estimated_delivery values are positive, i.e., orders are delivered later than expected.
3. Some Orders Delivered Early which are diff_estimated_delivery values < 0, indicating early deliveries.

Recommendations:

1. Geo-based Delivery Optimization is Explorring which geographies have longest delays And Prioritize partnership with quicker local carriers.
2. Set Realistic Estimated Delivery Dates:
   Modify order_estimated_delivery_date logic based on historical time_to_deliver and Better not to over-promise and under-deliver, particularly during peak seasons.
3. Check performance of logistics providers for delayed orders based on diff_estimated_delivery values < 0.

---

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

Answer:

Query:

```
with olc as (select round(avg(oi.freight_value),2) as state_freight,
       c.customer_state
       from `target.order_items` oi join `target.orders` o on oi.order_id=o.order_id
                          join `target.customers` c on o.customer_id=c.customer_id
        group by c.customer_state
),
  olv as (
    select customer_state,
           dense_rank()over(order by state_freight) as lowest_f,
           dense_rank() over(order by state_freight desc) as highest_f
           from olc
  )
select o.customer_state,
       o.state_freight from olc o join olv v on o.customer_state=v.customer_state
       where v.lowest_f <=5
union all
select o.customer_state,
       o.state_freight from olc o join olv v on o.customer_state=v.customer_state
       where v.highest_f <=5
```

Output:

Row	customer_state	state_freight
1	PB	42.72
2	PI	39.15
3	AC	40.07
4	RO	41.07
5	RR	42.98
6	SP	15.15
7	MG	20.63
8	DF	21.04
9	RJ	20.96
10	PR	20.53

Results per page: 50 1 – 10 of 10

Insights:

1. PB, PI, AC, RO, and RR states have the highest average freight amounts, all of which are above 39. This might be because they have greater distances from warehouses, lesser connectivity, or low volume of orders, resulting in cost per delivery.

2. SP, MG, DF, RJ, PR exhibit very low freight amounts, between 15–21.
   These are probably urban or logistics-optimized states with high order density and improved last-mile delivery infrastructures.

3. A gigantic gap of 28 from highest (RR: 42.98) to lowest (SP: 15.15) occurred due to chances of uneven balance of logistics cost-effectiveness in regions.

Recommendations:

1. Optimize High-Freight States:
   By Studying logistics partners, delivery routes, and warehousing opportunities in PB, RO, and RR we can Locally fulfill centers or bulk-shipping schemes can be planned.

2. Increase more marketing efforts in SP, MG, DF to further leverage low logistics cost.

3. Implement a state-wise freight normalization strategy — e.g., subsidizing high-cost states by modestly raising rates in low-cost states where margins are high.

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

Answer:

Query:

```
 with olc as ( select
round(avg(date_diff(o.order_delivered_customer_date,o.order_purchase_timestamp,day)),2) as
diff_in_delivery,

  c.customer_state
  from `target.orders` o
  join `target.customers` c on o.customer_id = c.customer_id
  group by c.customer_state
),
olv as (
  select customer_state,
         dense_rank() over(order by diff_in_delivery) as lowest_diff,
         dense_rank() over(order by diff_in_delivery desc) as highest_diff
  from olc
)
select *
from (
  select o.customer_state,
         o.diff_in_delivery
  from olc o
  join olv v on o.customer_state = v.customer_state
  where v.lowest_diff <= 5

  union all

  select o.customer_state,
         o.diff_in_delivery
  from olc o
  join olv v on o.customer_state = v.customer_state
  where v.highest_diff <= 5
) as final_result
order by diff_in_delivery desc;
```

Output:

Row	customer_state	diff_in_delivery
1	RR	28.98
2	AP	26.73
3	AM	25.99
4	AL	24.04
5	PA	23.32
6	SC	14.48
7	DF	12.51
8	MG	11.54
9	PR	11.53
10	SP	8.3

Results per page: 50 1 – 10 of 10

Insights:

1. Long Delays in Northern & Northeastern States:
   These states such as RR (Roraima), AP (Amapá), and AM (Amazonas) possess longest delivery periods (over 25 days) based probably on geographic remoteness, inadequate logistics facilities, or few shipments.

2. Southeastern States Are Fastest:
   SP (São Paulo), PR (Paraná), and MG (Minas Gerais) present significantly shorter delivery times (below 12 days) as an indicator of well-established logistics platforms and easy access to main hubs of distribution.

3. High Variance Across Regions:
   The difference between the fastest (SP at 8.3 days) and slowest (RR at 28.98 days) states is more than 20 days, reflecting high inconsistencies in delivery performance across states.

Recommendations:

1. Optimize Logistics in High-Delay States:
   Prioritize logistics in states such as RR, AP, and AM—collaborate with local couriers or establish regional warehouses to minimize shipping time and expense.

2. Analyze Distribution Strategies in Low-Delay States:
   Learn from best-performing states such as SP and MG—analytics of their warehousing, last-mile delivery model, carrier partnerships to emulate in poor-performing regions.

3. Install Regional SLA Targets:
   Establish State-level delivery SLAs and track them on a weekly basis. Utilize this information to nudge courier services to achieve performance targets and reward top-performing regions.

4. 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.

   Answer:

   Query:

```
select c.customer_state,

       round(avg(date_diff(o.order_estimated_delivery_date,
                           o.order_delivered_customer_date, day)), 2)
       as avg_early_delivery
       from  `target.orders` o join  `target.customers` c
                               on    o.customer_id = c.customer_id

       group by  c.customer_state
       order by avg_early_delivery desc
       limit 5;
```

   Output:

Row	customer_state	avg_early_delivery
1	AC	19.76
2	RO	19.13
3	AP	18.73
4	AM	18.61
5	RR	16.41

Results per page: 50 1 – 5 of 5

Insights:

1. AC and RO are outstanding with almost 20 days early delivery, way ahead of the mean. This indicates an extremely aggressive supply chain or a small order volume with rapid fulfillment.

2. All 5 states (AC, RO, AP, AM, RR) are from sparsely populated or geographically distant areas, which means possibly fewer traffic snarls, or high-priority dispatch owing to logistical practices.

3. The large early delivery gap implies that estimated delivery dates are likely padded or conservative, particularly in these areas. There may be a disconnect between expected vs actual performance of logistics partners.

Recommendations:

1. If deliveries are invariably 15–20 days in advance, the system can overestimate delivery timelines. Update the estimated delivery date model to project actual delivery capability for better customer trust and experience.

2. Look at implementing these states (AC, RO, etc.) as hub zones to fulfill priority or same-day delivery products locally, as infrastructure appears to be supportive.

3. Leverage the insight to promote fast delivery guarantee to customers in these markets it can enhance conversions and build brand reputation.

6.Analysis based on the payments:

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

Query:

```
select format_timestamp("%Y-%m", o.order_purchase_timestamp) as month,
       p.payment_type,
       count(*) as orders_placed
       from `target.orders` o join `target.payments` p on o.order_id=p.order_id
       group by 1,2
       order by month
```

Output:

Row	month	payment_type	orders_placed
1	2016-09	credit_card	3
2	2016-10	voucher	23
3	2016-10	credit_card	254
4	2016-10	UPI	63
5	2016-10	debit_card	2
6	2016-12	credit_card	1
7	2017-01	credit_card	583
8	2017-01	UPI	197
9	2017-01	debit_card	9
10	2017-01	voucher	61

Results per page: 50 1 – 50 of 90

Insights:

1. In 2017-11, 5897 orders placed through credit_card which is highest may be November sales offered by banks with related discounts.

2. There are not-defined value in payment type.
   may be it is Cash on delivery

3. Orders placed through
   credit card>UPI>voucher>debitcard>not defined

Recommendations:

1. Collaborate with banks to push credit card deals during busy months (such as festive or end-of-year sales)

2. Highlight EMI, cashback, or reward points in-app or onsite for increased conversions.

3. Define or Normalize Unspecified Payment Categories
   The occurrence of "not-defined" payment types indicates data inconsistency or unmonitored payment modes such as Cash on Delivery (COD).

4. Encourage UPI and voucher payments with fast checkout flows, special offers for a limited time, or loyalty points to minimize transaction fees over cards.

5. Offer "preferred payment discounts" or reflect estimated delivery enhancements for utilizing speedier payment instruments such as UPI.

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

Answer:

Query:

```
SELECT payment_installments,
       COUNT(DISTINCT order_id) AS num_orders
       FROM target.payments`
       WHERE payment_installments > 0
       GROUP BY payment_installments
       ORDER BY payment_installments;
```

Output:

Row	payment_installment	num_orders
1	1	49060
2	2	12389
3	3	10443
4	4	7088
5	5	5234
6	6	3916
7	7	1623
8	8	4253
9	9	644
10	10	5315

Results per page: 50 1 – 23 of 23

Insights:

1. Most Orders are Paid in One Installment:
   49,060 out of total orders (~54%) were paid in 1 installment, showing a high demand for full payment upfront.This may be due to customer confidence, or very high affordability for the product/service.

2. A sharp decline is observed from 1 to 2 installments (from 49,060 → 12,389), but orders are still fairly high for up to 6 installments.
   These short-term flexible options appear to be attractive to customers who want to stretch payments without long-term obligation.

3. Orders with more than 10 installments drop significantly, and most values fall below 150.
   Indicates that longer payment options are not frequently in demand, perhaps because interest charges may be higher, customers are hesitant, or less promoted.

Recommendation:

1. Market 1 to 6 Installment Plans
   Because most customers prefer shorter terms, emphasize flexible ones such as "Pay in 3" or "Pay in 6" in your promotional efforts. Give incentives for upfront payment to speed up cash flow.

2. Review and Revamp 10+ Installment Promises
   Low use indicates these lengthy plans might not be appealing or needed. Revise or delete them unless related to high-priced purchases.

3. Introduce Loyalty Rewards for Payments in Full