```
# Q1-State by Customers


WITH states AS
  (SELECT
    customer_state,
    COUNT(DISTINCT customer_unique_id )AS num_cust
  FROM target.customers
  GROUP BY customer_state),

rank_cust AS
  (SELECT
  *,
  ROW_NUMBER() OVER(ORDER BY num_cust DESC) rn
  FROM states)

SELECT
  customer_state,
  num_cust,
  CONCAT (ROUND((SUM (num_cust) OVER(ORDER BY num_cust DESC) /SUM (num_cust)
OVER())*100,2),'%')
  AS    cum_num_of_cust
FROM rank_cust
ORDER BY num_cust DESC

# Q2-Pareto product category by orders



SELECT
  num_ord,
  product_category,
  CONCAT(ROUND(SUM(num_ord) OVER (ORDER BY rn)/SUM(num_ord) OVER()*100,2),'%') AS
  cum_ord_perc,
  CONCAT(ROUND(COUNT(product_category) OVER (ORDER By rn)/COUNT(product_category)
  OVER()*100,2),'%') AS cum_cat_perc
FROM  (SELECT
        *,
        ROW_NUMBER () OVER(ORDER BY num_ord DESC) rn
      FROM (SELECT
              COUNT(DISTINCT o.order_id) num_ord,
              p.product_category
            FROM target.orders o
            LEFT JOIN target.order_items oi
            ON o.order_id = oi.order_id
            LEFT JOIN target.products p
            ON oi.product_id=p.product_id
            GROUP BY p.product_category)ord_table
      )rn_table
ORDER BY rn
```

```
# Q3-Pareto product id by orders



SELECT
  num_ord,
  product_id,
  CONCAT(ROUND(SUM(num_ord) OVER (ORDER BY rn)/SUM(num_ord) OVER()*100,4),'%') AS
cum_ord,
  CONCAT(ROUND(COUNT(product_id) OVER (ORDER By rn)/COUNT(product_id)
OVER()*100,2),'%') AS cum_prod
FROM (SELECT
        *,
        ROW_NUMBER () OVER(ORDER BY num_ord DESC) rn
      FROM (SELECT
              COUNT(DISTINCT o.order_id) num_ord,
              p.product_id
            FROM target.orders o
            LEFT JOIN target.order_items oi
            ON o.order_id = oi.order_id
            LEFT JOIN target.products p
            ON oi.product_id=p.product_id
            GROUP BY p.product_id)cnt_ord_table
      )rn_table
ORDER BY rn;


#4- Quandrant chart


SELECT
  ROUND(AVG(o.review_score),2) avg_review,
  COUNT(DISTINCT oi.order_id) num_ord,
  p.product_category
FROM target.order_reviews o
LEFT JOIN target.order_items oi
ON o.order_id = oi.order_id
LEFT JOIN target.products p
ON oi.product_id = p.product_id
GROUP BY product_category
HAVING num_ord >=1000 AND avg_review >=4.10
ORDER BY avg_review DESC;

SELECT
  COUNT(DISTINCT product_id) total_product
FROM target.products

SELECT
  COUNT(DISTINCT product_category) total_product
FROM target.products
```

```
#target data


#1 number of customers


SELECT
  COUNT(DISTINCT customer_unique_id) num_customers
FROM target.customers;


#2 Customer by Year,month

SELECT
  *,
  SUM(num_cust) OVER (ORDER BY year,month) AS cum_cust
FROM (SELECT
        EXTRACT(YEAR FROM o.order_purchase_timestamp) year,
        EXTRACT(MONTH FROM o.order_purchase_timestamp) month,
        COUNT(DISTINCT c.customer_unique_id) num_cust
      FROM target.customers c
      LEFT JOIN target.orders o
      ON c.customer_id=o.customer_id
      GROUP BY  year,month
      ORDER BY year,month)cust_table;



#3 Orders

SELECT
  *,
  SUM(ord) OVER (ORDER BY year,month) AS cum_ord
FROM (SELECT
        EXTRACT(YEAR FROM order_purchase_timestamp) year,
        EXTRACT(MONTH FROM order_purchase_timestamp) month,
        COUNT(DISTINCT order_id) ord
      FROM target.orders o
      GROUP BY  year,month
      ORDER BY year,month)order_table;



#4 Net promoter Score


SELECT
```

```
  CONCAT(ROUND((((SUM(IF(review_score = 5, 1, 0)))-(SUM(IF(review_score = 1, 1,
0)) ))/
  COUNT(review_score))*100,2),'%') net_promoter_score
FROM target.order_reviews;


#5 RATING BY ORDERS


SELECT
  review_score,
  num_review,
  CONCAT(ROUND((num_review/ SUM(num_review) OVER())*100,2),'%') percentage
FROM (SELECT
        review_score,
        COUNT(review_score) num_review,
      FROM target.order_reviews
      GROUP BY review_score)review_table
ORDER BY review_score;


#6 AVRG DURATION BY REVIEW


SELECT
  DISTINCT review_score,
  ROUND(AVG(day) OVER(PARTITION BY review_score),2) AS avg_approval_day,
  ROUND(AVG(day) OVER(),2) AS avg_approval_day_overall
FROM (SELECT
        TIMESTAMP_DIFF(o.order_approved_at,o.order_purchase_timestamp,SECOND)/86400
day,
        ov.review_score
      FROM target.orders o
      RIGHT JOIN target.order_reviews ov
      ON o.order_id=ov.order_id
      WHERE o.order_approved_at IS NOT NULL)day_table
ORDER BY review_score



#7 AVRG DURATION CARRIER PICKUP


SELECT
  DISTINCT review_score,
  ROUND(AVG(day) OVER(PARTITION BY review_score),2) AS avg_carrier_day,
  ROUND(AVG(day) OVER(),2) AS overall_avg_carrier_day
FROM (SELECT

TIMESTAMP_DIFF(o.order_delivered_carrier_date,o.order_approved_at ,SECOND)/86400 day,
        ov.review_score
      FROM target.orders o
      RIGHT JOIN target.order_reviews ov
      ON o.order_id=ov.order_id
      WHERE o.order_approved_at IS NOT NULL)day_table
ORDER BY review_score


#8 No. of Orders by Status & Avg. Duration Estimated vs Delivery


SELECT
  DISTINCT o.order_status,
  ov.review_score,
  COUNT(DISTINCT ov.order_id) OVER (PARTITION BY o.order_status,ov.review_score)
num_ord,
  ROUND(AVG(COALESCE
(TIMESTAMP_DIFF(o.order_delivered_customer_date,o.order_estimated_delivery_date ,SECON
D)/86400,0))
  OVER (PARTITION BY o.order_status),2)day_avg
FROM target.orders o
RIGHT JOIN target.order_reviews ov
ON o.order_id=ov.order_id
ORDER BY o.order_status,ov.review_score;




#9 customers by sates and review score

SELECT
  DISTINCT c.customer_state,
  COUNT(ov.review_id)OVER(PARTITION BY c.customer_state)num_reviews,
  ov.review_score,
  CONCAT(ROUND((COUNT (ov.review_score)
  OVER (PARTITION BY c.customer_state,ov.review_score)/
  COUNT(ov.order_id) OVER (PARTITION BY c.customer_state))*100,2),'%') rev_percentage,
  COUNT(ov.order_id) OVER (PARTITION BY c.customer_state) num_review_by_state
FROM target.customers c
LEFT JOIN target.orders o
ON c.customer_id=o.customer_id
LEFT JOIN target.order_reviews ov
ON o.order_id=ov.order_id
WHERE review_score IS NOT NULL
ORDER BY c.customer_state,ov.review_score




#10 orders


SELECT
  COUNT(DISTINCT order_id)num_ord
```

```sql
FROM target.orders


#11 AVG order per seller


SELECT
  ROUND((COUNT (DISTINCT o.order_id))/
  (COUNT(DISTINCT s.seller_id)),2) avg_ord_per_seller
FROM target.orders o
LEFT JOIN target.order_items i
ON o.order_id=i.order_id
LEFT JOIN  target.sellers s
ON s.seller_id=i.seller_id;




#12 AVG DURATION OF estmated


SELECT

ROUND((SUM(TIMESTAMP_DIFF(order_estimated_delivery_date,order_purchase_timestamp,SECON
D)/86400 ))/
  (COUNT( order_estimated_delivery_date)),2) avg_duration_esimated
FROM target.orders


#13 TOTAL SELLERS

SELECT
  COUNT(DISTINCT seller_id) sellers
FROM target.sellers

#14 NO OF ORDERS BY PAYMENT INSTALLMENT

SELECT
  payment_installments,
  COUNT (order_id) num_ord
FROM target.payments
GROUP BY payment_installments
ORDER BY payment_installments




#15 AVG DURATION OF DELIVERY BY STATE

SELECT
  DISTINCT c.customer_state,
ROUND(AVG(TIMESTAMP_DIFF(o.order_estimated_delivery_date,o.order_purchase_timestamp,SE
COND)/86400)
  OVER(PARTITION BY c.customer_state),2) avg_day_delivery
FROM target.orders o
LEFT JOIN target.customers c
ON o.customer_id=c.customer_id

#16 AVG DURATION BY MONTH AND YEAR



SELECT
  DISTINCT year,
  month,
  ROUND(AVG(day) OVER (PARTITION BY year,month),2) AS avg_duration
FROM (SELECT
        EXTRACT(YEAR FROM order_purchase_timestamp) year,
        EXTRACT(MONTH FROM order_purchase_timestamp) month,

TIMESTAMP_DIFF(o.order_estimated_delivery_date,o.order_purchase_timestamp,SECOND)/8640
0 day
      FROM target.orders o)day_table
ORDER BY year,month



#17 HOUR BY PAYMENT METHOD COUND ORDER

SELECT
  DISTINCT day,
  payment_type,
  month,
  COUNT(order_id) OVER (PARTITION BY day,month,payment_type) num_ord
FROM (SELECT
        CASE  WHEN (EXTRACT (HOUR FROM o.order_purchase_timestamp)) <=6 THEN 'Dawn'
              WHEN (EXTRACT (HOUR FROM o.order_purchase_timestamp)) <=12 THEN
'Morning'
              WHEN (EXTRACT (HOUR FROM o.order_purchase_timestamp)) <=18 THEN
'Afternoon'
              WHEN (EXTRACT (HOUR FROM o.order_purchase_timestamp)) <=23 THEN'Night'
END day,
        p.payment_type,
        EXTRACT (MONTH FROM o.order_purchase_timestamp) month,
        o.order_id
      FROM target.payments p
      LEFT JOIN target.orders o
      ON p.order_id=o.order_id
      WHERE o.order_purchase_timestamp IS NOT NULL AND p.payment_type IS NOT NULL)t
ORDER BY CASE
  WHEN day = 'Dawn' THEN 1
  WHEN day = 'Morning' THEN 2
  WHEN day = 'Afternoon' THEN 3
  WHEN day = 'Night' THEN 4
  END,payment_type,month;
```

```
#18 TOP 5 STATES BY DELIVERY DURATION


SELECT
  DISTINCT c.customer_state,

ROUND(AVG(TIMESTAMP_DIFF(o.order_delivered_customer_date,o.order_purchase_timestamp,SE
COND)/86400)
  OVER(PARTITION BY c.customer_state),2) avg_delivery_day
FROM target.orders o
LEFT JOIN target.customers c
ON o.customer_id=c.customer_id
ORDER BY avg_delivery_day DESC
LIMIT 5

#19 state by deliver day minus estimated


SELECT
  customer_state,
  avg_day_dele_esti
FROM (SELECT
        c.customer_state,

ROUND(AVG(TIMESTAMP_DIFF(o.order_delivered_customer_date,o.order_estimated_delivery_da
te,SECOND)/86400),2) avg_day_dele_esti,
        DENSE_RANK()
        OVER (ORDER BY ROUND(AVG(TIMESTAMP_DIFF(o.order_delivered_customer_date,
o.order_estimated_delivery_date, SECOND) / 86400), 2)) AS rn
      FROM target.orders o
      LEFT JOIN target.customers c
      ON o.customer_id=c.customer_id
      GROUP BY c.customer_state)ranked
WHERE rn <= 5
ORDER BY rn;

#20   TOP 5 STATE BY FRIEGHT VALUE



SELECT
  c.customer_state,
  ROUND(AVG(oi.freight_value ),2) avg_fr
FROM  target.customers c
RIGHT JOIN target.orders o
ON o.customer_id=c.customer_id
RIGHT JOIN target.order_items oi
ON o.order_id=oi.order_id
```

```
GROUP BY c.customer_state
ORDER BY avg_fr DESC
LIMIT 5;


#21 5 SLOWEST DURATION


SELECT
  c.customer_state,

ROUND(AVG(TIMESTAMP_DIFF(o.order_delivered_customer_date,o.order_purchase_timestamp,SE
COND)/86400),2) avg_delivery_day
FROM target.orders o
LEFT JOIN target.customers c
ON o.customer_id=c.customer_id
GROUP BY c.customer_state
ORDER BY avg_delivery_day
LIMIT 5;

# 22 AVG ORDER VALUE


SELECT
  ROUND(SUM(payment_value)/COUNT(DISTINCT order_id),2) avg_order_value
FROM target.payments


#23 TOTAL REVENUE


SELECT
  ROUND(SUM(payment_value)) AS total_revenue
FROM target.payments;


#24 AVG freight value

SELECT
  ROUND(SUM(freight_value)/COUNT(DISTINCT order_id),2) avg_freight_value
FROM target.order_items


#25 total revenue by  month

SELECT
  FORMAT_TIMESTAMP('%b %Y', order_purchase_timestamp) ord_date,
  ROUND(SUM(p.payment_value),2) revenue
FROM target.payments p
LEFT JOIN target.orders o
```

```sql
ON p.order_id=o.order_id
GROUP BY ord_date
ORDER BY PARSE_DATE('%b %Y', ord_date)
```