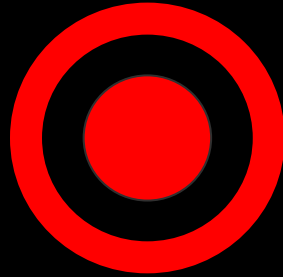


CASE STUDY OF TARGET



USING SQL AND TABLEAU

1. No. of unique cities and states :

QUERY :

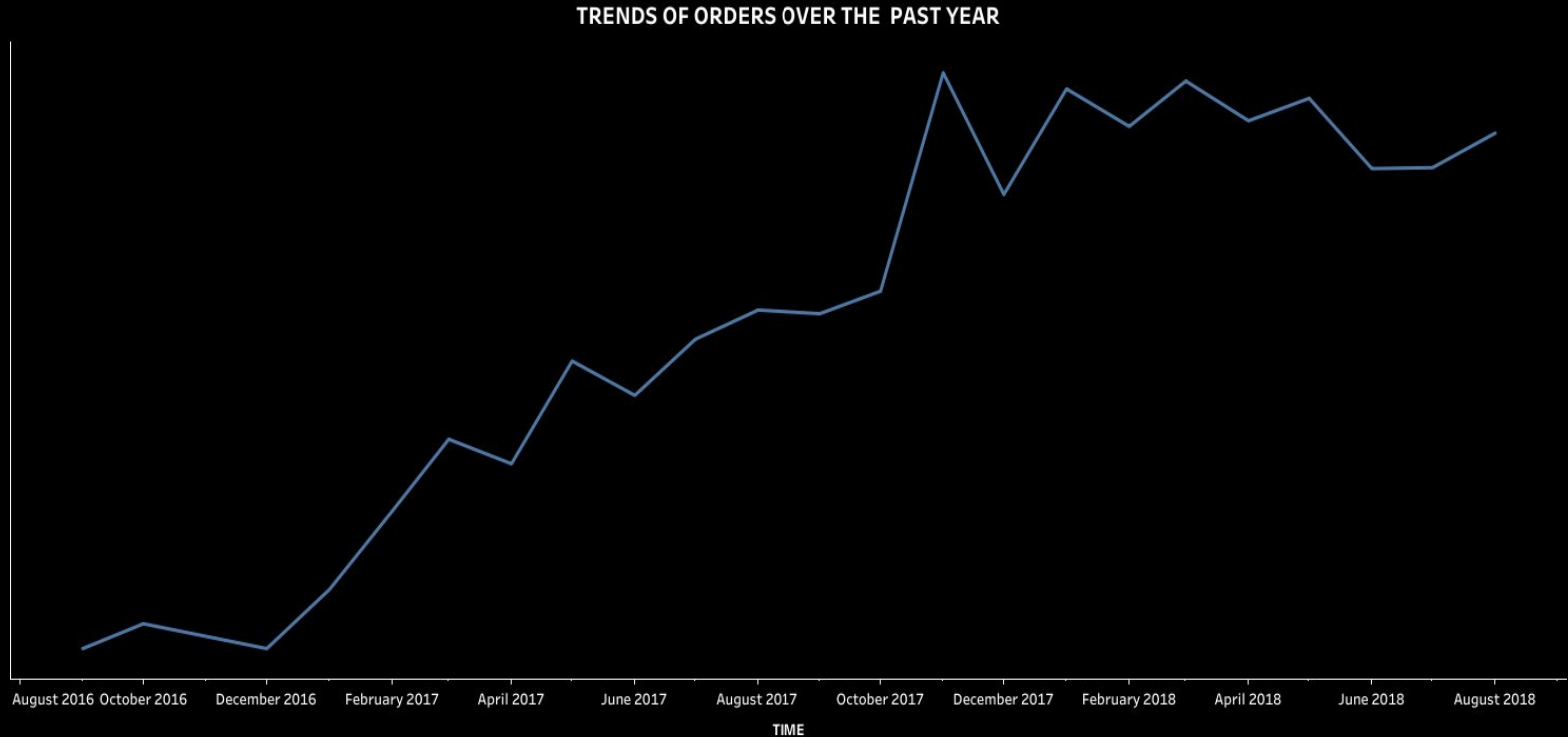
```
SELECT  
    COUNT(DISTINCT geolocation_city) AS NO_OF_CITIES,  
    COUNT(DISTINCT geolocation_state) AS NO_OF_STATES  
FROM `project.geolocation`
```

OUTPUT :

NO_OF_CITIES ▼	NO_OF_STATES ▼
8011	27

From the result of the query,
we can see that there are 8011 cities
and 27 different states

2. Is there a growing trend in the numbers of orders over the past years?



3. Top 10 category with maximum number of orders

QUERY:

```
SELECT  
UPPER(product_category) AS Category,  
COUNT(*) AS No_of_orders  
FROM `project.order_items` AS o  
JOIN `project.products` AS p  
USING(product_id)  
GROUP BY product_category  
ORDER BY 2 DESC  
LIMIT 10
```

OUTPUT:

Category ▼	No_of_orders ▼
BED TABLE BATH	11115
HEALTH BEAUTY	9670
SPORT LEISURE	8641
FURNITURE DECORATION	8334
COMPUTER ACCESSORIES	7827
HOUSEWARES	6964
WATCHES PRESENT	5991
TELEPHONY	4545
GARDEN TOOLS	4347
AUTOMOTIVE	4235

TOP 10 MOST ORDERED PRODUCT CATEGORIES

BED TABLE BATH



HEALTH BEAUTY



SPORT LEISURE



FURNITURE
DECORATION



COMPUTER
ACCESSORIES



HOUSEWARES



WATCHES PRESENT



TELEPHONY



GARDEN TOOLS



AUTOMOTIVE



4. No. of orders placed during Dawn, Mornings, Afternoon and Night.

QUERY :

```
WITH tab_a AS (  
  SELECT order_id,  
  CASE WHEN EXTRACT(HOUR FROM order_purchase_timestamp) BETWEEN 0 AND 6  
  THEN 'Dawn'  
  WHEN EXTRACT(HOUR FROM order_purchase_timestamp) BETWEEN 7 AND 12  
  THEN 'Morning'  
  WHEN EXTRACT(HOUR FROM order_purchase_timestamp) BETWEEN 13 AND 18  
  THEN 'Afternoon'  
  WHEN EXTRACT(HOUR FROM order_purchase_timestamp) BETWEEN 19 AND 23  
  THEN 'Night'  
  END AS TIME_OF_DAY  
  FROM project.orders  
)  
SELECT  
  TIME_OF_DAY,  
  COUNT(*) AS NO_OF_ORDERS  
  FROM tab_a  
  GROUP BY TIME_OF_DAY
```

OUTPUT:

TIME_OF_DAY ▼	NO_OF_ORDERS ▼
Morning	27733
Dawn	5242
Afternoon	38135
Night	28331

Here, we can see that most of orders were made during afternoon (13-18) and good numbers of orders were also made during Morning (7-12) and Night (19-23). Less number of orders were placed during Dawn hours (0-6).

Now, let's see what kind of products were placed the most during separate times of day.

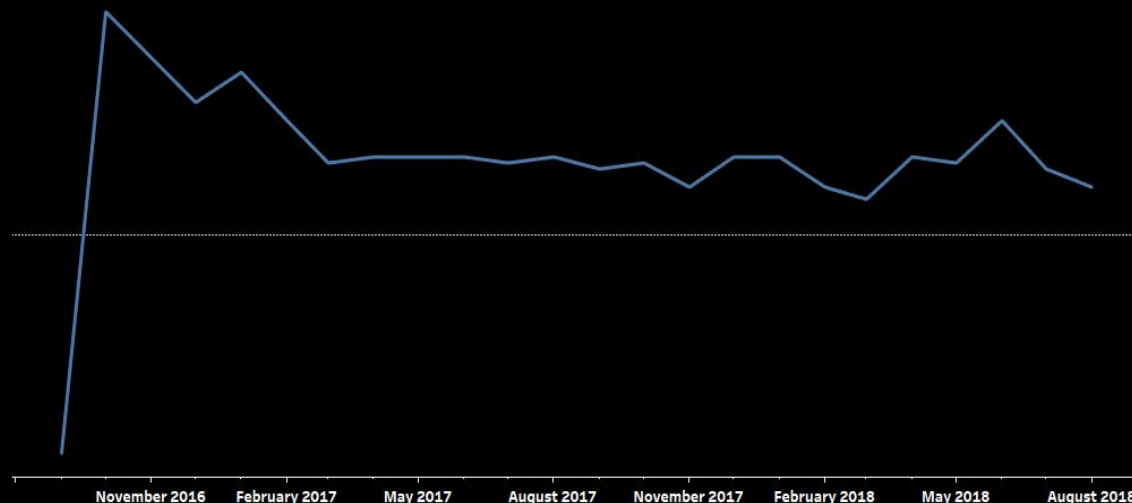
TOP 3 PRODUCTS ORDERED FOR EACH TIME OF DAY:

QUERY:

```
WITH tab_a AS (  
  SELECT product_category,  
    CASE WHEN EXTRACT(HOUR FROM order_purchase_timestamp) BETWEEN 0 AND 6  
    THEN 'Dawn'  
    WHEN EXTRACT(HOUR FROM order_purchase_timestamp) BETWEEN 7 AND 12  
    THEN 'Morning'  
    WHEN EXTRACT(HOUR FROM order_purchase_timestamp) BETWEEN 13 AND 18  
    THEN 'Afternoon'  
    WHEN EXTRACT(HOUR FROM order_purchase_timestamp) BETWEEN 19 AND 23  
    THEN 'Night'  
  END AS TIME_OF_DAY  
FROM 'project.orders' AS o JOIN 'project.order_items' AS ot  
ON o.order_id = ot.order_id JOIN 'project.products' AS p  
ON ot.product_id = p.product_id  
  SELECT TIME_OF_DAY, product_category, COUNT(*) AS ct  
FROM tab_a  
GROUP BY TIME_OF_DAY, product_category  
  SELECT TIME_OF_DAY, UPPER(product_category) AS CATEGORY,  
    DENSE_RANK() OVER(PARTITION BY TIME_OF_DAY ORDER BY ct DESC) AS rank  
FROM tab_b  
) SELECT * FROM tab_c  
WHERE rank <=3
```


OUTPUT :

mt	AVERAGE_DELAYED_DELIVERY_TJ
2016-9	-36.0
2016-10	37.0
2016-12	22.0
2017-1	27.0
2017-2	19.0
2017-3	12.0
2017-4	13.0
2017-5	13.0
2017-6	13.0
2017-7	12.0
2017-8	13.0
2017-9	11.0
2017-10	12.0
2017-11	8.0



We can see a gradually decreased in the delivery time taken from 2016 to 2017 and 2018. But there was a spike going upwards in May 2018 and it decreased afterwards.

OUTPUT:

TIME_OF_DAY ▼	CATEGORY ▼	rank ▼
Morning	BED TABLE BATH	1
Morning	HEALTH BEAUTY	2
Morning	COMPUTER ACCESSORIES	3
Afternoon	BED TABLE BATH	1
Afternoon	HEALTH BEAUTY	2
Afternoon	SPORT LEISURE	3
Night	BED TABLE BATH	1
Night	HEALTH BEAUTY	2
Night	FURNITURE DECORATION	3
Dawn	BED TABLE BATH	1
Dawn	HEALTH BEAUTY	2
Dawn	SPORT LEISURE	3

From the output, we can say that BED TABLE BATH and HEALTH BEAUTY category is the most ordered category regardless of the day time. COMPUTER ACCESSORIES were placed mostly during Morning and FURNITURE AND DECORATION during night, and SPORT LEISURE during Dawn.

5. Top 10 states with maximum number of customers.

QUERY:

```
SELECT  
customer_state,  
COUNT(DISTINCT customer_id) AS NO_OF_CUSTOMERS  
FROM project.customers`  
GROUP BY customer_state  
ORDER BY 2 DESC
```

OUTPUT:

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

customer_state ▼	NO_OF_CUSTOMERS
RR	46
AP	68
AC	81
AM	148
RO	253
TO	280
SE	350
AL	413
RN	485
PI	495

Most of customers are from Sao Paulo (SP), Rio de Janeiro (RJ) and Minas Gerais (MG). Whereas, Roraima (RR), Amapa (AP) and Acre (AC) have the least numbers of customers. And rest of the states are in between.

Viz. of distribution of customers state-wise :



6. Average difference between the Estimated & Actual delivery date of each month

QUERY:

```
WITH tab_a AS (  
  SELECT  
    EXTRACT(YEAR FROM order_purchase_timestamp) AS YEAR,  
    EXTRACT(MONTH FROM order_purchase_timestamp) AS MONTH,  
    TIMESTAMP_DIFF(EXTRACT(DATE FROM order_estimated_delivery_date), EXTRACT(DATE FROM  
order_delivered_customer_date), DAY) AS ag  
  FROM `project.orders`  
) SELECT  
  CONCAT(YEAR, '-', MONTH) AS mt,  
  ROUND(AVG(ag), 0) AS AVERAGE_DELAYED_DELIVERY_TIME  
  FROM tab_a  
  GROUP BY YEAR, MONTH  
  ORDER BY YEAR, MONTH
```

7. Most used payments method.

QUERY :

```
SELECT  
payment_type,  
COUNT(order_id) AS no_of_orders  
FROM project.payments`  
GROUP BY payment_type  
ORDER BY 2 DESC
```

OUTPUT :

payment_type ▼	no_of_orders ▼
credit_card	76795
UPI	19784
voucher	5775
debit_card	1529

PIE CHART OF PAYMENT TYPES USED :

