

# TARGET

**1) Importing the dataset and do usual exploratory analysis steps like checking the structure & characteristics of the dataset**

- 1. Data type of columns in a table**
- 2. Time period for which the data is given**
- 3. Cities and States of customers ordered during the given period**

## Data type of columns in a table

### CUSTOMERS

```
select column_name ,data_type from information_schema.columns
```

where table\_name="customers";

COLUMN_NAME	DATA_TYPE
customer_city	varchar
customer_id	varchar
customer_state	char
customer_unique_id	varchar
customer_zip_code_prefix	int

### GEOLOCATION

```
select column_name ,data_type from information_schema.columns  
where table_name="geolocation";
```

COLUMN_NAME	DATA_TYPE
geolocation_lat	float
geolocation_lng	float
geolocation_city	varchar

geolocation_state	char
geolocation_zip_code_prefix	int
geolocation_id	int

## ORDER\_REVIEWS

```
select column_name ,data_type from information_schema.columns
where table_name="order_reviews"
```

COLUMN_NAME	DATA_TYPE
review_id	varchar
order_id	varchar
review_score	int
review_comment_title	varchar
review_comment_message	varchar
review_creation_date	timestamp
review_answer_timestamp	timestamp

## ORDER\_ITEMS

```
select column_name ,data_type from information_schema.columns
where table_name="order_items";
```

COLUMN_NAME	DATA_TYPE
order_id	varchar
order_item_id	int
product_id	varchar
seller_id	varchar
shipping_limit_date	timestamp
price	float
freight_value	float

## ORDERS

```
select column_name ,data_type from information_schema.columns
```

```
where table_name="orders";
```

COLUMN_NAME	DATA_TYPE
order_id	varchar
customer_id	varchar
order_status	varchar
order_purchase_timestamp	timestamp
order_approved_at	timestamp
order_delivered_carrier_date	timestamp
order_delivered_customer_date	timestamp
order_estimated_delivery_date	timestamp

## PAYMENTS

```
select column_name ,data_type from information_schema.columns  
where table_name="payments";
```

COLUMN_NAME	DATA_TYPE
order_id	varchar
payment_sequential	int
payment_type	varchar
payment_installments	int
payment_value	float

## PRODUCTS

```
select column_name ,data_type from information_schema.columns  
where table_name="products";
```

COLUMN_NAME	DATA_TYPE
product_id	varchar
product_category_name	varchar
product_name_lenght	int
product_description_lenght	int

product_photos_qty	int
product_weight_g	int
product_length_cm	int
product_height_cm	int
product_width_cm	int

## SELLERS

```
select column_name ,data_type from information_schema.columns
where table_name="sellers";
```

COLUMN_NAME	DATA_TYPE
seller_city	varchar
seller_id	varchar
seller_state	char
seller_zip_code_prefix	int

## Time period for which the data is given

```
Select min(order_purchase_timestamp) as start_date,
max(order_purchase_timestamp) as end_date
from orders;
```

start_date	end_date
04-09-2016 21:15	17-10-2018 17:30

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

```
Select count(distinct customer_city) AS distinct_city_count,
count(distinct customer_state) AS distinct_state_count
from customers
```

```
Join orders on customers.customer_id = orders.customer_id
where order_purchase_timestamp BETWEEN '2016-09-04' AND '2018-10-17';
```

distinct_city_count	distinct_state_count
---------------------	----------------------

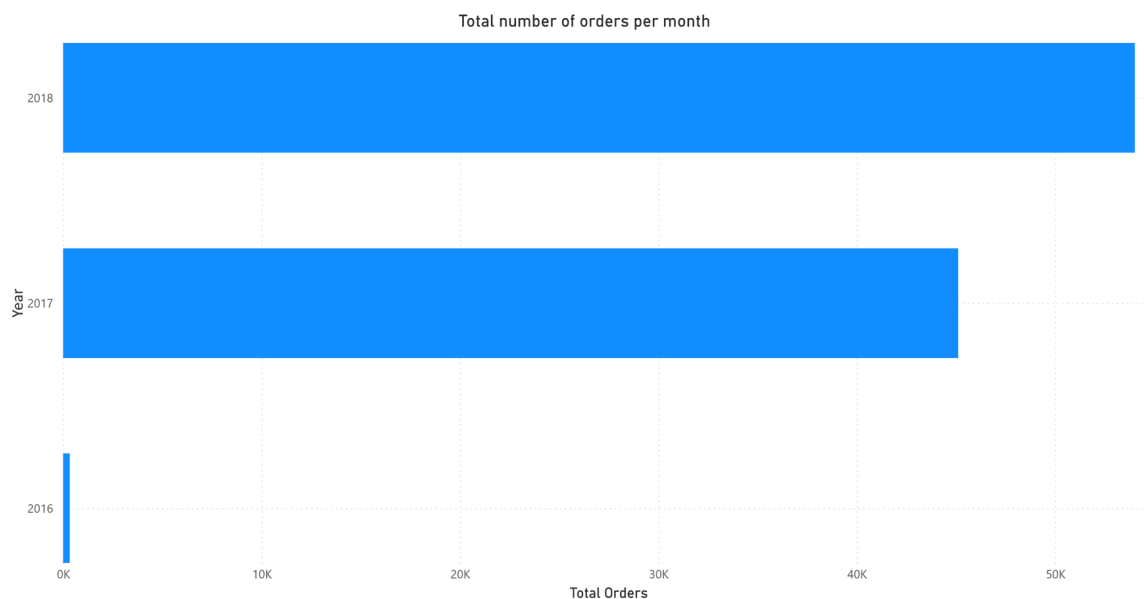
4119	27
------	----

## 2)In-depth Exploration:

- Is there a growing trend on e-commerce in Brazil?
- How can we describe a complete scenario?
- Can we see some seasonality with peaks at specific months?
- What time do Brazilian customers tend to buy (Dawn, Morning, Afternoon or Night)
- Is there a growing trend on e-commerce in Brazil?
- How can we describe a complete scenario?
- Can we see some seasonality with peaks at specific months?

## Total number of orders per month

```
Select date_format (order_purchase_timestamp, '%Y-%m') as mon
count(*) as num_orders from orders group by month order by mo
```

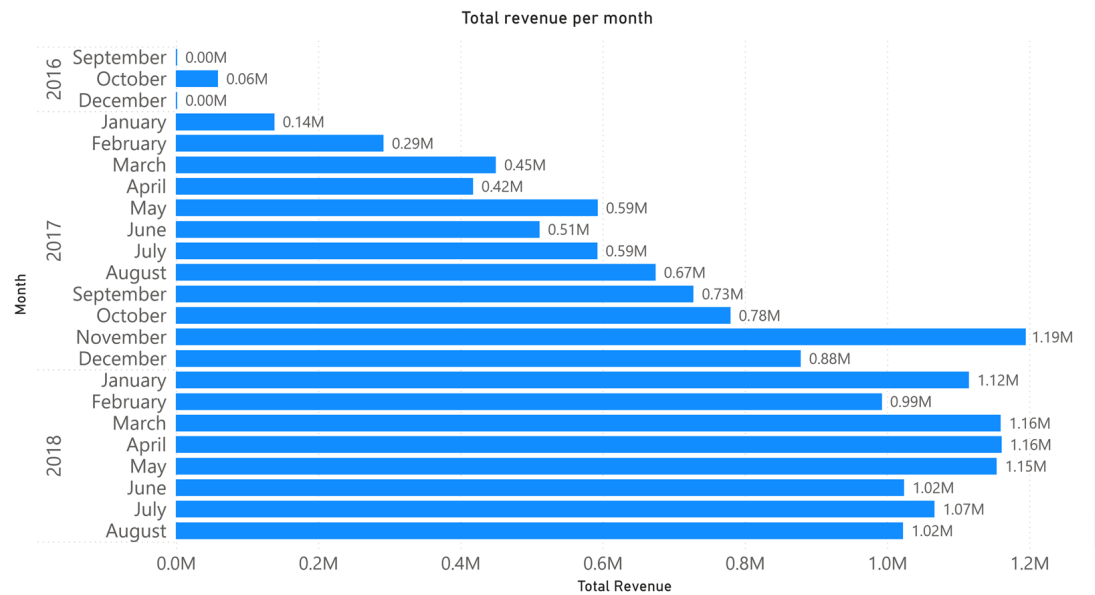


## Total revenue per month

```

Select date_format(order_purchase_timestamp, '%Y-%m') as month
Sum(payment_value) as total_revenue
From orders o
Join payments p on o.order_id = p.order_id
Group by month
Order BY month;

```

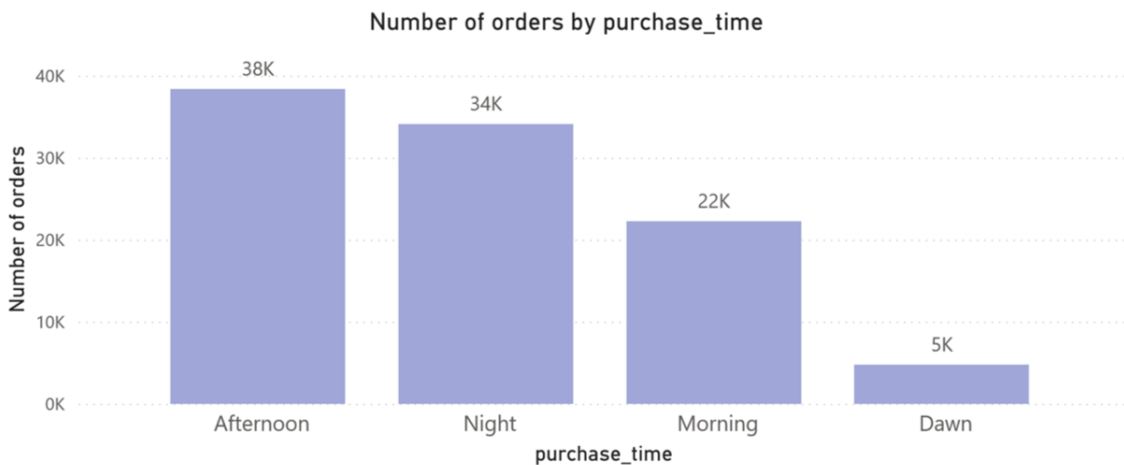


### What time do Brazilian customers tend to buy (Dawn, Morning, Afternoon or Night)?

```

select case
when hour (order_purchase_timestamp) >= 0 and hour(order_purc
then 'Dawn' when hour (order_purchase_timestamp) >= 6 and hou
then
'Morning' when hour (order_purchase_timestamp) >= 12 and hour
'Afternoon'
else 'Night' end as purchase_time, count(*) AS num_orders from
group by purchase_time
limit 0, 10;

```

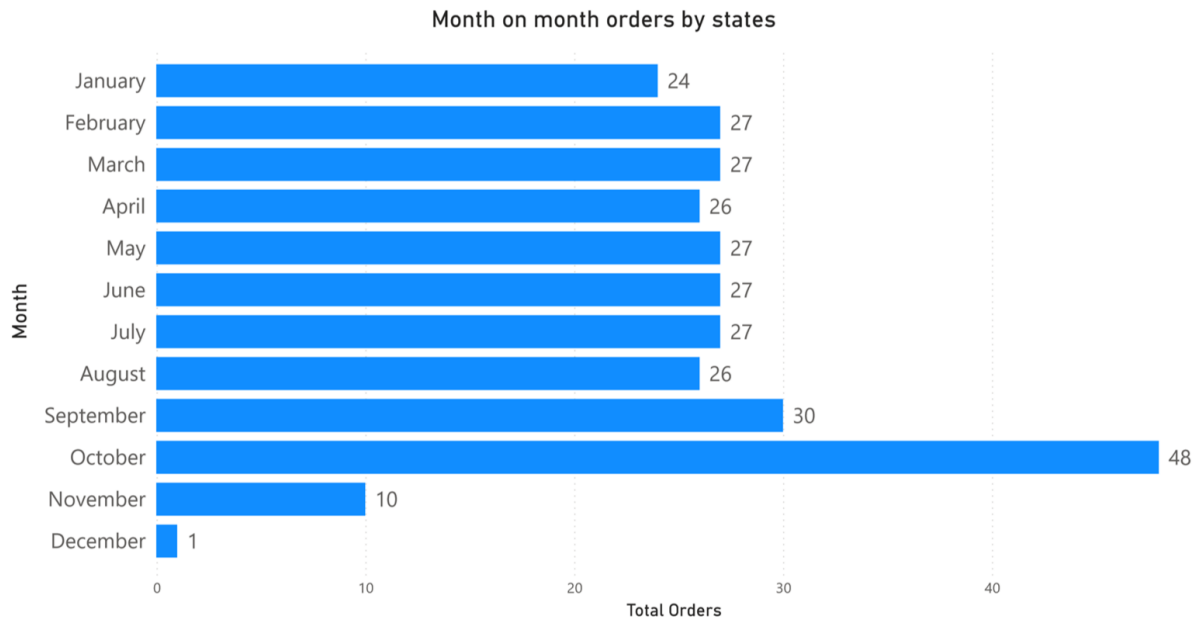


### 3) Evolution of E-commerce orders in the Brazil region:

1. Get month on month orders by states
2. Distribution of customers across the states in Brazil

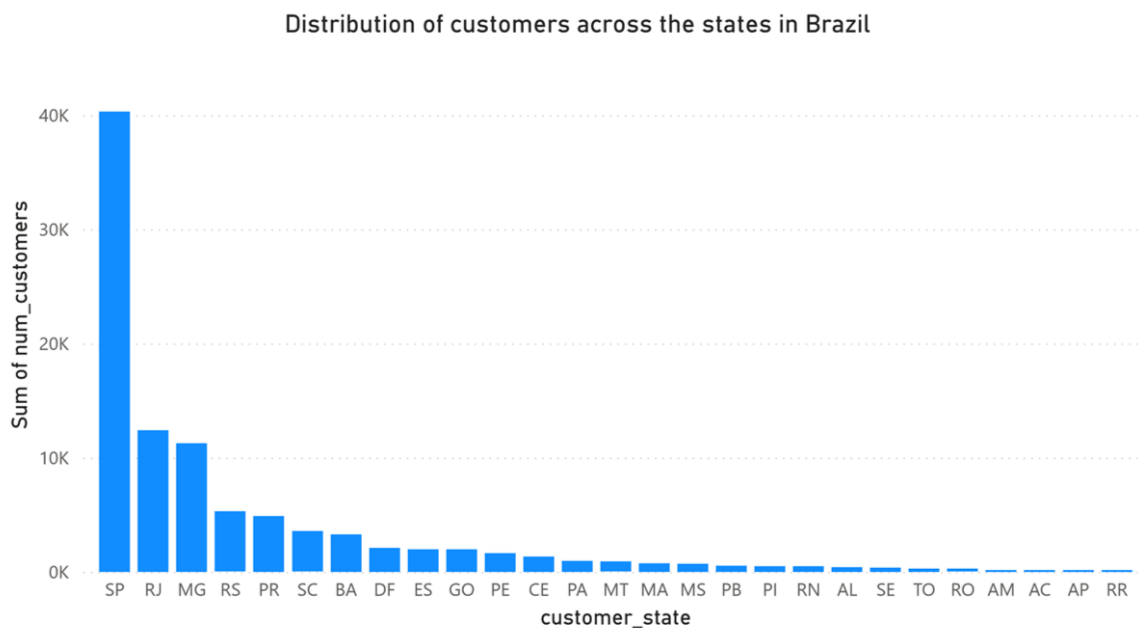
#### Get month on month orders by states

```
select date_format (order_purchase_timestamp, '%Y-%m') AS mon
count(*) AS num_orders from orders o
join customers c ON o.customer_id = c.customer_id
group by month, customer_state
order by month, customer_state;
```



### Distribution of customers across the states in Brazil

```
select customer_state, count(distinct customer_unique_id) AS
from customers group by customer_state
order by num_customers desc;
```





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

1) Mean & Sum of price and freight value by customer state

##### **Mean & Sum of price and freight value by customer state**

```
select c.customer_state, round(avg(op.price), 2) AS avg_price,
round(sum(op.price), 2) AS sum_price,
round(avg(op.freight_value), 2) AS avg_freight,
round(sum(op.freight_value), 2) AS sum_freight from order_items op
join orders o on op.order_id = o.order_id
join customers c on o.customer_id = c.customer_id
where o.order_status in ('delivered', 'approved') group by c.customer_state
```

customer_state	avg_price	sum_price	avg_freight	sum_freight
SP	109.1	5067803.06	15.12	702079.55
RS	118.83	728897.47	21.61	132575.32
SC	123.75	507012.13	21.51	88115.65
BA	134.02	493584.14	26.49	97553.67
MS	142.33	115429.97	23.35	18937.58
RJ	124.42	1759651.13	20.91	295750.44
PI	161.99	84721	39.12	20457.19
MG	120.18	1552521.53	20.62	266431.76
ES	120.74	268643.45	22.03	49014.48
RO	167.34	45682.76	41.33	11283.24

#### **5) Analysis on sales, freight and delivery time**

Find time\_to\_delivery & diff\_estimated\_delivery

Group data by state, take mean of freight\_value, time\_to\_delivery, diff\_estimated\_delivery

Sort the data to get the following:

Top 5 states with highest/lowest average freight value - sort in desc/asc limit 5

Top 5 states with highest/lowest average time to delivery

Top 5 states where delivery is really fast/ not so fast compared to estimated date

time\_to\_delivery = order\_purchase\_timestamp - order\_delivered\_customer\_date

### Find time\_to\_delivery & diff\_estimated\_delivery

```
select o.order_id,
```

```
timestampdiff(day, o.order_purchase_timestamp, o.order_delivered_customer_date)
as time_to_delivery,
timestampdiff(day, o.order_estimated_delivery_date, o.order_delivered_customer_date)
as diff_estimated_delivery from orders o
where o.order_status = 'delivered';
```

order_id	time_to_delivery	diff_estimated_delivery
e481f51cbdc54678b7cc49136f2d6af7	8	-7
53cdb2fc8bc7dce0b6741e2150273451	13	-5
47770eb9100c2d0c44946d9cf07ec65d	9	-17
949d5b44dbf5de918fe9c16f97b45f8a	13	-12
ad21c59c0840e6cb83a9ceb5573f8159	2	-9
a4591c265e18cb1dcee52889e2d8acc3	16	-5
6514b8ad8028c9f2cc2374ded245783f	9	-11
76c6e866289321a7c93b82b54852dc33	9	-31
e69bfb5eb88e0ed6a785585b27e16dbf	18	-6
e6ce16cb79ec1d90b1da9085a6118aeb	12	-8

### Grouping data by state, take mean of freight\_value, time\_to\_delivery, diff\_estimated\_delivery

```
select c.customer_state, avg(oi.freight_value) as mean_freight_value,
avg(timestampdiff(DAY, o.order_purchase_timestamp,
o.order_delivered_customer_date)) as mean_time_to_delivery,
avg(timestampdiff(DAY, o.order_estimated_delivery_date,
o.order_delivered_customer_date)) as mean_diff_estimated_delivery
```

```
from orders o join customers c on o.customer_id = c.customer_id join
order_items oi on o.order_id = oi.order_id
group by c.customer_state;
```

### **Sorting the data to get the following:**

```
select c.customer_state, avg(oi.freight_value) as mean_freight_value,
avg(timestampdiff(day, o.order_purchase_timestamp,
o.order_delivered_customer_date)) as mean_time_to_delivery,
AVG(timestampdiff(day, o.order_estimated_delivery_date,
o.order_delivered_customer_date)) as mean_diff_estimated_delivery
from orders o join customers c on o.customer_id = c.customer_id
join order_items oi on o.order_id = oi.order_id
group by c.customer_state
order by mean_freight_value desc, mean_time_to_delivery asc;
```

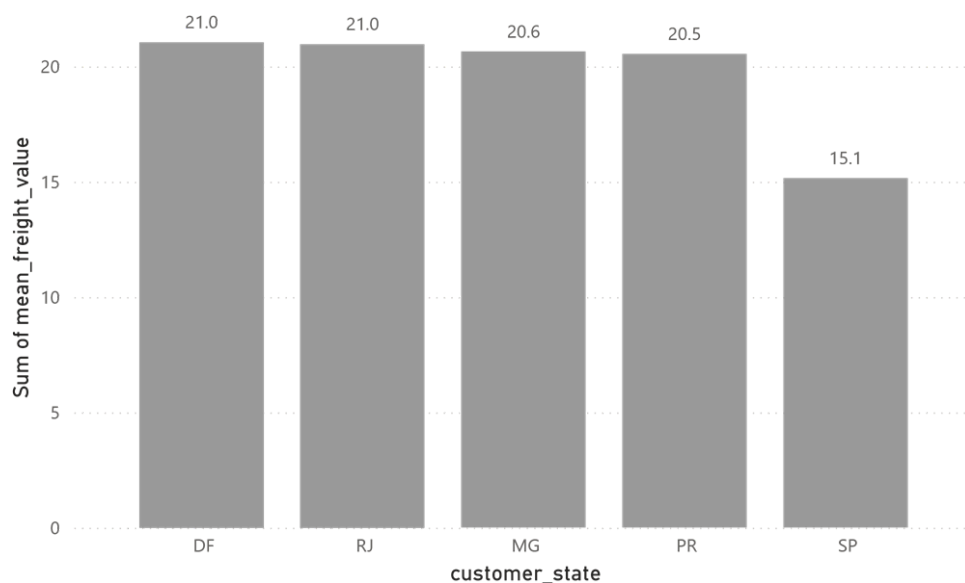
### **Top 5 states with highest/lowest average freight value - sort in desc/asc limit 5**

```
select c.customer_state, avg(oi.freight_value) as mean_freigh
join customers c on o.customer_id = c.customer_id
join order_items oi on o.order_id = oi.order_id
group by c.customer_state
order by mean_freight_value DESC
limit 5;
```

customer_state	mean_freight_value
RR	42.98442307692309
PB	42.723803986711
RO	41.06971223021582
AC	40.0733695652174
PI	39.147970479704824

### lowest average freight value

```
select c.customer_state, avg(oi.freight_value) as mean_freigh
join customers c on o.customer_id = c.customer_id
join order_items oi on o.order_id = oi.order_id
group by c.customer_state
order by mean_freight_value asc
limit 5;
```



### Top 5 states with highest/lowest average time to delivery

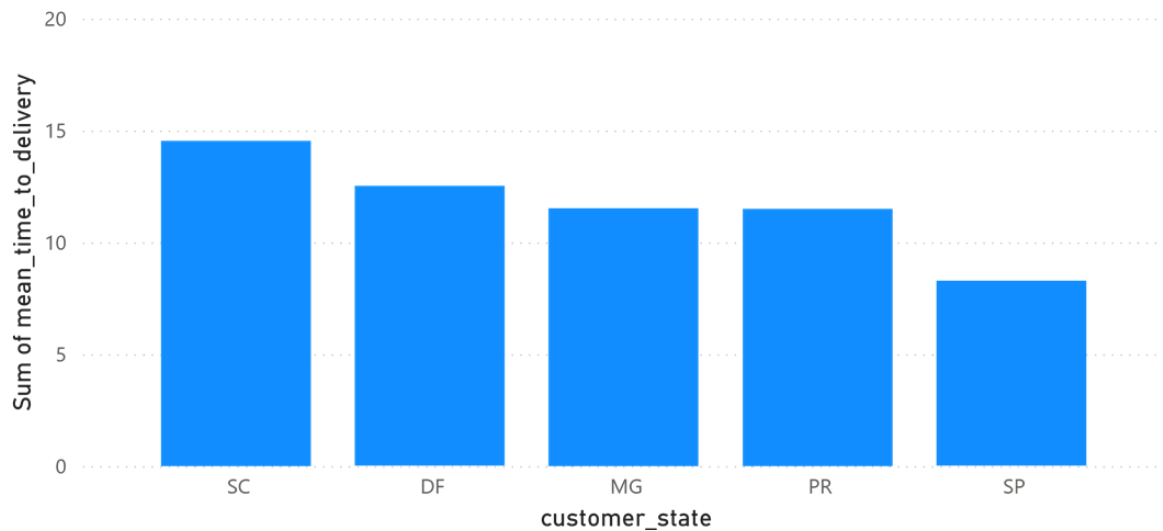
```
select c.customer_state,  
avg(timestampdiff(day, o.order_purchase_timestamp, o.order_delivered_customer_date))  
AS mean_time_to_delivery  
from orders o  
join customers c on o.customer_id = c.customer_id  
join order_items oi on o.order_id = oi.order_id  
group by c.customer_state  
order by mean_time_to_delivery desc  
limit 5;
```



### lowest average time to delivery

```
select c.customer_state, avg(timestampdiff(day, o.order_purchase_timestamp,  
o.order_delivered_customer_date))  
AS mean_time_to_delivery from orders o  
join customers c on o.customer_id = c.customer_id  
join order_items oi on o.order_id = oi.order_id
```

```
group by c.customer_state
order by mean_time_to_delivery asc
limit 5;
```



### **Top 5 states where delivery is really fast/ not so fast compared to estimated date**

```
Select c.customer_state,
avg(timestampdiff(day, o.order_estimated_delivery_date,
o.order_delivered_customer_date)) as mean_diff_estimated_delivery
from orders o
join customers c on o.customer_id = c.customer_id
join order_items oi on o.order_id = oi.order_id
group by c.customer_state having mean_diff_estimated_delivery < 0
order BY mean_diff_estimated_delivery ASC
limit 5;
```

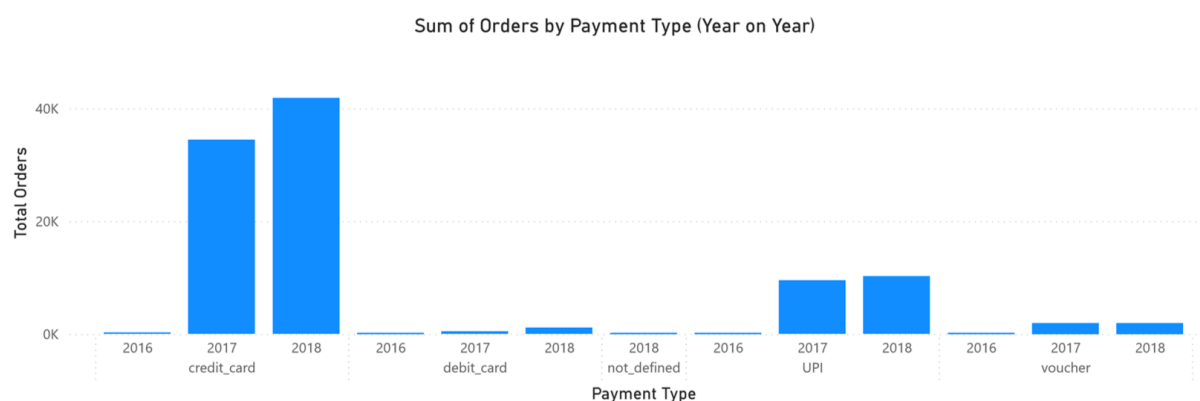
customer_state	mean_diff_estimated_delivery
AC	-20.0110
RO	-19.0806
AM	-18.9755
AP	-17.4444
RR	-17.4348

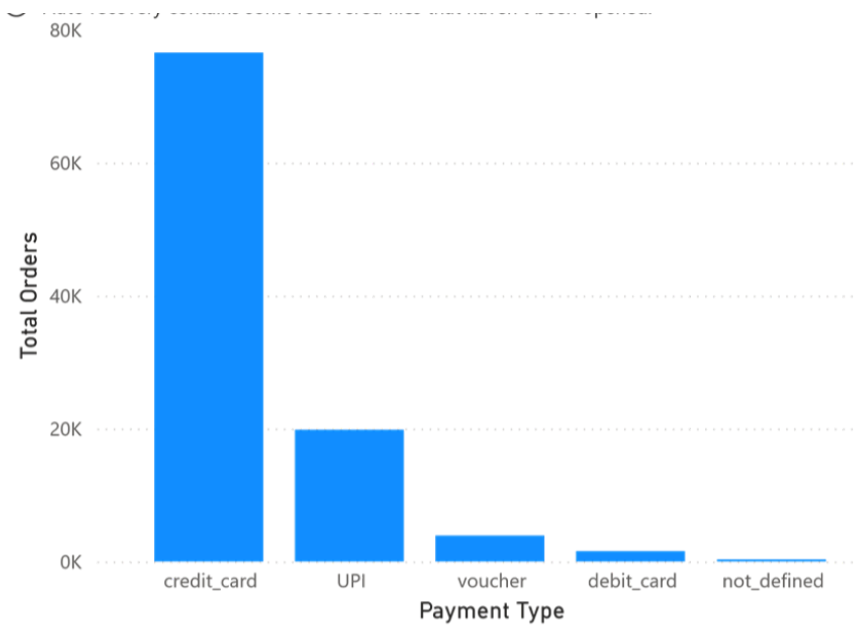
## 6)Payment type analysis:

- year on year count of orders for different payment types
- Count of orders based on the no. of payment instalments

### year on year count of orders for different payment types

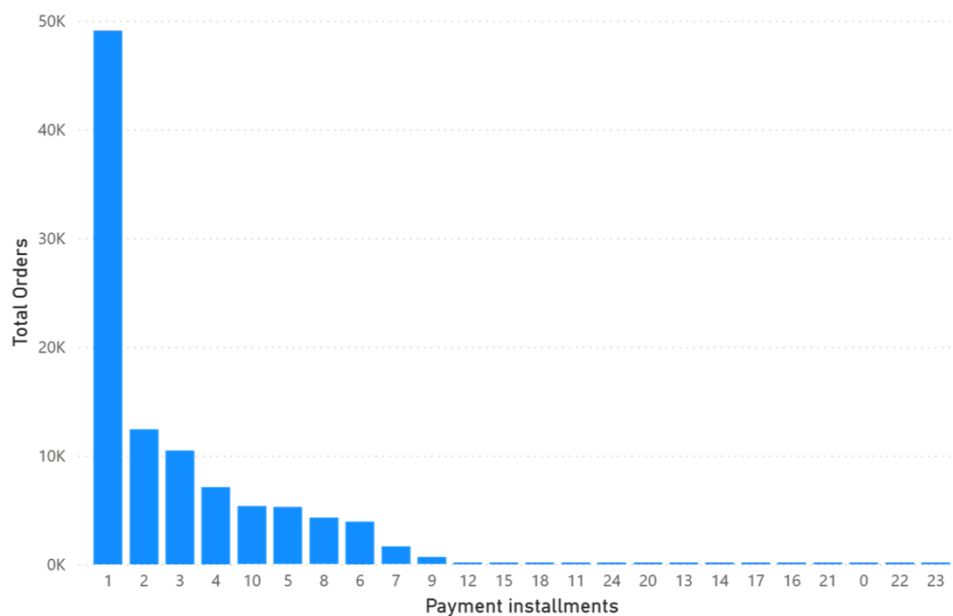
```
select Year(o.order_purchase_timestamp) as year,
month(o.order_purchase_timestamp) as month,
p.payment_type, count(distinct o.order_id) as order_count from
join orders o on p.order_id = o.order_id
group by year(o.order_purchase_timestamp),
month(o.order_purchase_timestamp), p.payment_type;
```





### Count of orders based on the no. of payment instalments

```
select p.payment_installments, count(distinct o.order_id) as o
from payments p
join orders o on p.order_id = o.order_id
group by p.payment_installments
```





## **Key Findings:**

### **1. Time Period and Data Overview:**

- The data spans from September 4, 2016, to October 17, 2018.
- There are 4119 distinct cities and 27 states from which customers placed orders during this period.

### **2. E-commerce Trends:**

- There is a growing trend in e-commerce orders over time in Brazil.
- Monthly order counts and revenue show variations, indicating possible seasonality.
- Analysis of purchase times shows distinct trends with orders placed during dawn, morning, afternoon, and night.

### **3. Customer Distribution:**

- The distribution of customers varies significantly across different states in Brazil.
- A detailed breakdown by state reveals differences in the number of unique customers.

### **4. Economic Impact:**

- There is significant variation in mean and total order prices and freight values across different states.
- Some states have higher average order values and freight costs compared to others.

### **5. Delivery Analysis:**

- The average delivery time and the difference between estimated and actual delivery dates vary by state.
- Identified the top states with the highest and lowest average freight values and delivery times.

### **6. Payment Methods:**

- The count of orders varies based on payment types and the number of installments.
- Year-on-year analysis of payment methods shows trends and preferences among customers.

## **Recommendations:**

### **1. Growth and Seasonality:**

- Focus marketing and promotional efforts during peak months identified in the seasonal analysis.
- Use the insights from purchase time trends to optimize advertising and sales strategies throughout the day.

### **2. State-Specific Strategies:**

- Tailor strategies to target states with higher customer distributions and order values.
- For states with lower average order values, consider localized promotions to boost sales.

### **3. Delivery Optimization:**

- Work on improving delivery times in states with higher than average delivery durations.
- Investigate the reasons behind the discrepancies between estimated and actual delivery dates to enhance customer satisfaction.

### **4. Freight Cost Management:**

- Negotiate better shipping rates or explore alternative logistics partners for states with high freight costs.
- Offer incentives for bulk orders in states with lower freight costs to maximize profitability.

### **5. Payment Method Diversification:**

- Promote payment methods that are preferred by customers to increase convenience and potentially boost sales.
- Consider offering flexible installment options, especially in regions where this payment method is popular.

### **6. Further Analysis:**

- Conduct deeper analysis into the factors influencing order values and delivery times, such as product categories and seller performance.
- Regularly update the analysis to track changes and adjust strategies accordingly.

These recommendations can help in optimizing operations, enhancing customer satisfaction, and driving growth in the e-commerce space.