

1. Exploratory Data Analysis:

1.1 Data type of all columns in the "customers" table.

Query:

```
SELECT column_name, data_type
FROM Abc_superstore.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

Insight:

From the above output, we can see that there are 5 columns and 2 different data types in the "customers" table.

1.2 Time range between which the orders were placed.

Query:

```
SELECT
  MIN(order_purchase_timestamp) AS minimum,
  MAX(order_purchase_timestamp) AS maximum,
  DATE_DIFF(MAX(order_purchase_timestamp), MIN(order_purchase_timestamp), DAY) AS
time_range_days
FROM
  Abc_superstore.orders;
```

Output:

Row	minimum	maximum	time_range_days
1	2016-09-04 21:15:19 UTC	2018-10-17 17:30:18 UTC	772

Insight:

We can see that the operations of Abc_superstore in Brazil began on 4th Sept 2016 and the last order placed in this dataset is on 17th Oct 2018, which is 772 days apart.

1.3 Cities & States of customers who ordered during the given period

Query:

```
SELECT
  COUNT(DISTINCT(geolocation city)) AS num_cities,
  COUNT(DISTINCT(geolocation_state)) AS num_states
FROM
  Abc_superstore.geolocation;
```

Output:

Row	num_cities	num_states
1	8011	27

Insight:

Customers of Abc_superstore in Brazil are spread across 8011 cities from 27 states.

2. In-depth Exploration:

2.1 Trend in the no. of orders placed over the past years

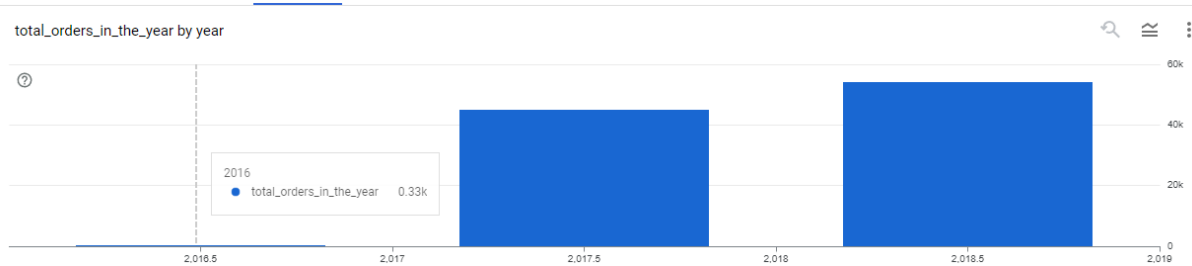
Query:

```
SELECT * FROM
(
  SELECT EXTRACT(year FROM order_purchase_timestamp) AS year,
  COUNT(order_id) AS total_orders_in_the_year
  FROM Abc_superstore.orders
  GROUP BY EXTRACT(year FROM order_purchase_timestamp)
) t
ORDER BY year;
```

Output:

Row	year	total_orders_in_the_year
1	2016	329
2	2017	45101
3	2018	54011

Insight:



Since operations started on Sept 2016, we see less orders placed on that year. But we can see a significant rise in the number of orders placed with Abc_superstore Brazil over the years. It shows that Abc_superstore is a brand liked by the Brazilians.

2.2 Monthly seasonality in terms of the number of orders placed:

Query :

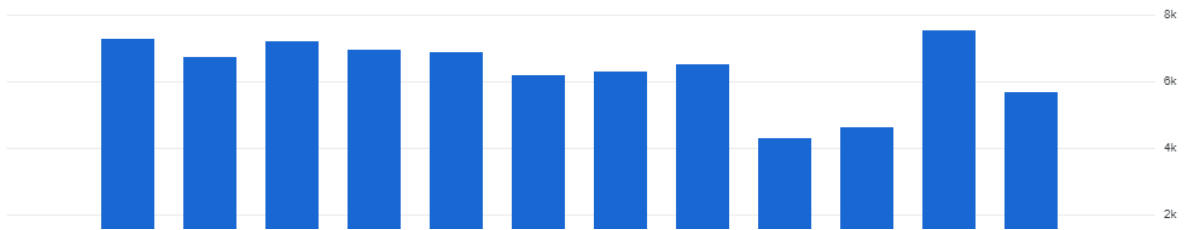
```
WITH cte as(
    SELECT EXTRACT(month FROM order_purchase_timestamp) AS month,
           COUNT(order_id) AS total_orders_in_the_month,
           EXTRACT(year FROM order_purchase_timestamp) AS year
    FROM Abc_superstore.orders
    GROUP BY EXTRACT(month FROM order_purchase_timestamp),
             EXTRACT(year FROM order_purchase_timestamp)
)
select month,year, total_orders_in_the_month, sum(total_orders_in_the_month) over()
as grand_total
from cte
order by year,month;
```

Sample output :

Row	month ▼	year ▼	total_orders_in_the_r	grand_total ▼
1	9	2016	4	99441
2	10	2016	324	99441
3	12	2016	1	99441
4	1	2017	800	99441
5	2	2017	1780	99441
6	3	2017	2682	99441
7	4	2017	2404	99441
8	5	2017	3700	99441
9	6	2017	3245	99441
10	7	2017	4026	99441

Insight :

total_orders_in_the_month by month



From the above graph, we can conclude that the number of orders placed are seasonal. We can see that there is a dip in the orders placed during the months of September and October, and the highest sales happen during November.

Recommendation :

Products can be offered at a discounted price during the months of September and October.

2.3 Time of the day (Dawn, Morning, Afternoon or Night) during which the Brazilian customers placed the highest number of orders:

- 0-6 hrs : Dawn
- 7-12 hrs : Mornings
- 13-18 hrs : Afternoon
- 19-23 hrs : Night

Query:

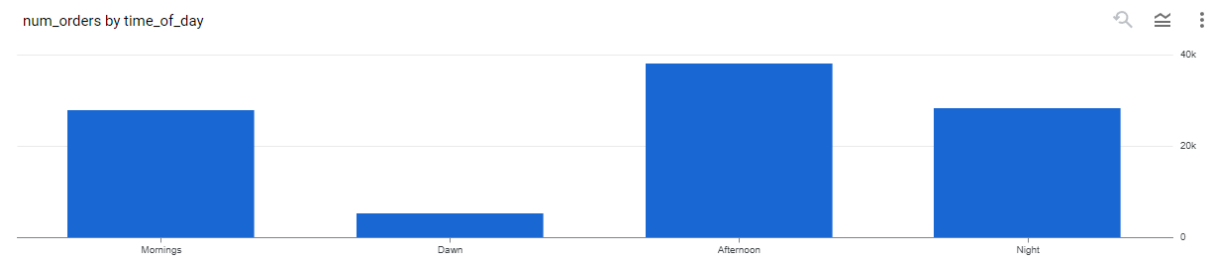
```
with cte as
(select
order_purchase_timestamp,
case when extract (hour from order_purchase_timestamp) BETWEEN 0 AND 6 THEN 'Dawn'
when extract (hour from order_purchase_timestamp) BETWEEN 7 AND 12 THEN 'Mornings'
when extract (hour from order_purchase_timestamp) BETWEEN 13 AND 18 THEN
'Afternoon'
when extract (hour from order_purchase_timestamp) BETWEEN 19 AND 23 THEN 'Night'
end as time_of_day
from Abc_superstore.orders)

select time_of_day, count(*) as num_orders from cte
group by time_of_day;
```

Output:

Row	time_of_day	num_orders
1	Mornings	27733
2	Dawn	5242
3	Afternoon	38135
4	Night	28331

Insight:



We can see that the highest number of orders were placed on Afternoons, followed by Nights and Mornings with almost equal number of orders. The least number of orders were placed during the Dawn.

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

3.1 Month on month no. of orders placed in each state:

Query:

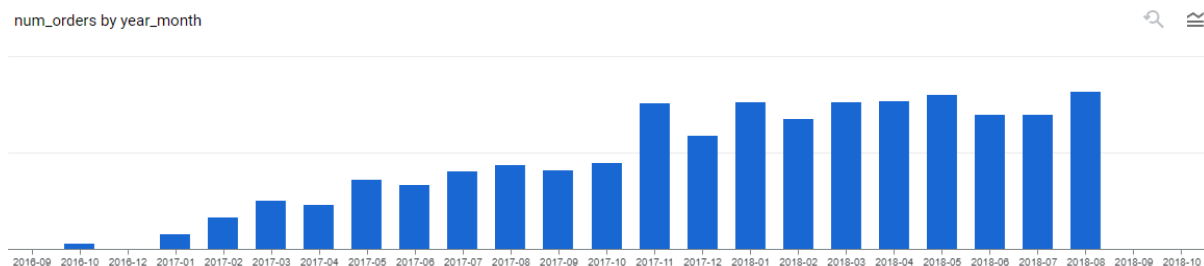
```
WITH cte as
(
SELECT distinct(o.order_id),
format_date('%Y-%m', order_purchase_timestamp) as year_month,
g.geolocation_state
FROM Abc_superstore.customers c
LEFT JOIN Abc_superstore.orders o ON
c.customer_id = o.customer_id
LEFT JOIN Abc_superstore.geolocation g ON
customer_zip_code_prefix = geolocation_zip_code_prefix
)

SELECT year_month, geolocation_state, count(order_id) as num_orders
FROM cte
GROUP BY year_month, geolocation_state
ORDER BY year_month, num_orders desc;
```

Sample output:

Row	year_month	geolocation_state	num_orders
1	2016-09	SP	2
2	2016-09	RR	1
3	2016-09	RS	1
4	2016-10	SP	113
5	2016-10	RJ	56
6	2016-10	MG	40
7	2016-10	RS	24
8	2016-10	PR	19
9	2016-10	SC	11
10	2016-10	GO	9
11	2016-10	CE	8
12	2016-10	PE	7

Insight:



From the graph above, we can conclude that the highest orders were placed on Nov in 2017, and on Jan, March, April, May, and Aug in 2018.

3.2 Customer distribution across all the states

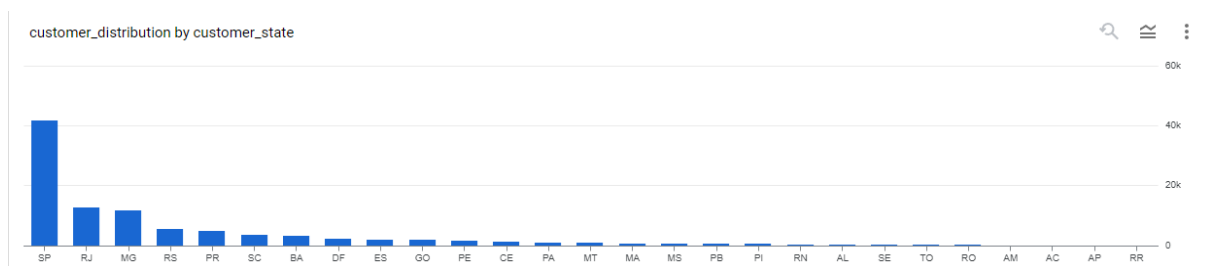
Query:

```
SELECT customer_state, count(customer_id) as customer_distribution,
FROM Abc_superstore.customers
GROUP BY customer_state
ORDER BY customer_distribution desc;
```

Sample Output:

Row	customer_state	customer_distribution
1	SP	41746
2	RJ	12852
3	MG	11635
4	RS	5466
5	PR	5045
6	SC	3637
7	BA	3380
8	DF	2140
9	ES	2033
10	GO	2020
11	PE	1652
12	CE	1336

Insight :



The state SP has the highest number of customers, with almost 3 times more than the second highest state – RJ, which is followed closely by the state MG. The 4th highest state, RS, has lesser than half the number of customers from MG.

- There is still a huge scope for growth in many other states.

Recommendation:

Marketing campaigns can be launched offering special discounts in the states with less number of customers.

4. Impact on Economy: Analysis of the money movement by e-commerce by looking at order prices, freight etc.

4.1 Percentage increase in the cost of orders from year 2017 to 2018 (for months between Jan to Aug only).

Query:

```
WITH cte as(
SELECT p.payment_value,
FORMAT_TIMESTAMP("%d/%m/%Y",o.order_purchase_timestamp) as formatted_ts,
sum(payment_value)over() as total_cost_17
FROM Abc_superstore.payments p
LEFT JOIN Abc_superstore.orders o ON
p.order_id=o.order_id
WHERE o.order_purchase_timestamp BETWEEN '2017-01-01' AND '2017-08-31'),

cte2 as(
SELECT p.payment_value,
FORMAT_TIMESTAMP("%d/%m/%Y",o.order_purchase_timestamp) as formatted_ts,
sum(payment_value)over() as total_cost_18
FROM Abc_superstore.payments p
LEFT JOIN Abc_superstore.orders o ON
p.order_id=o.order_id
WHERE o.order_purchase_timestamp BETWEEN '2018-01-01' AND '2018-08-31'
)

SELECT ROUND(((cte2.total_cost_18-cte.total_cost_17)/total_cost_17)*100,2) as
percentage_increase FROM cte, cte2
limit 1;
```

Output:

Row	percentage_increase
1	138.53

Insight:

From the above output, it is clear that there is an increase of **138.53%** in the cost of orders placed **between Jan-Aug, 2017 & 2018**.

4.2 Total & Average value of order price for each state.

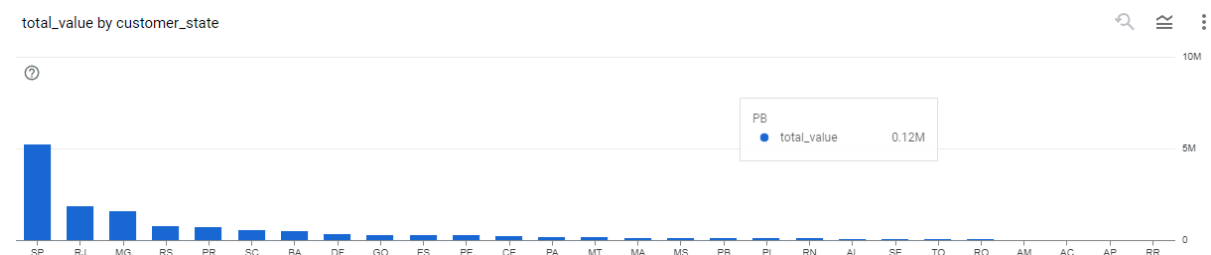
Query:

```
SELECT customer_state, total_value, avg_value
FROM (SELECT c.customer_state,
ROUND(sum(oi.price)over(partition by c.customer_state),2) as total_value,
ROUND(avg(oi.price)over(partition by c.customer_state),2) as avg_value
FROM Abc_superstore.customers c
LEFT JOIN Abc_superstore.orders o ON c.customer_id = o.customer_id
LEFT JOIN Abc_superstore.order_items oi ON o.order_id = oi.order_id) t
group by customer_state,total_value, avg_value
order by total_value desc;
```

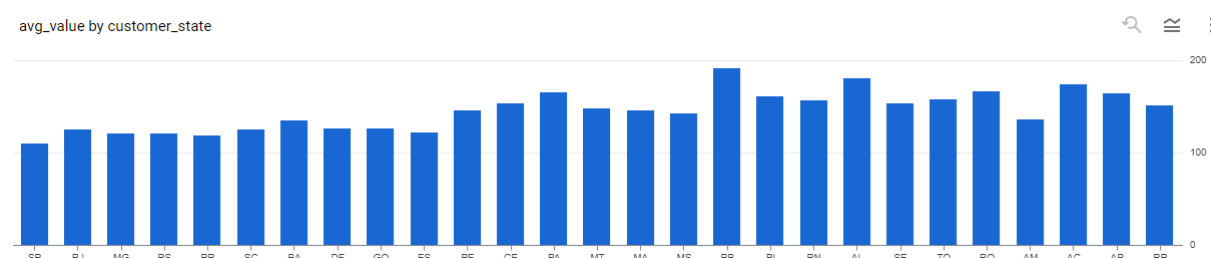

Sample output:

Row	customer_state	total_value	avg_value
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
10	ES	275037.31	121.91
11	PE	262788.03	145.51
12	CE	227254.71	153.76

Insight :



We can see that the state with the highest total order price is SP, which has the average value of order price equal to 109.65 Brazilian Reals.



The states PB, AL, PA, AC have high average order price values, even though their total value of order price is less. It may indicate that the states have **more affluent customers**. For example, the state with the 12th highest total order price, CE, has an average order price of 153.76 despite not making it to the top 10 states in total value.

Recommendation :

More luxury products can be sold in the states PB, AL, PA, AC, and CE where the average order price is higher than most of the other states.

4.3 Total & Average value of order freight for each state.

Query:

```
SELECT customer_state, total_value, avg_value
FROM (
SELECT c.customer_state,
ROUND(sum(oi.freight_value)over(partition by c.customer_state),2) as total_value,
ROUND(avg(oi.freight_value)over(partition by c.customer_state),2) as avg_value
FROM Abc_superstore.customers c
LEFT JOIN Abc_superstore.orders o ON c.customer_id = o.customer_id
LEFT JOIN Abc_superstore.order_items oi ON o.order_id = oi.order_id
)t
group by customer_state,total_value, avg_value
order by total_value desc;
```

Sample output:

Row	customer_state	total_value	avg_value
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

Insight:

SP has the highest total freight value but the lowest average freight value, indicating a high number of orders placed from that state, which lowers the average cost of freight charges.

5. Analysis based on sales, freight, and delivery time.

5.1 No. of days taken to deliver each order from the order's purchase date (as delivery time) and the difference (in days) between the estimated & actual delivery date of an order:

Query:

```
SELECT order_id, order_delivered_customer_date, order_purchase_timestamp,  
date_diff(order_delivered_customer_date,order_purchase_timestamp,day) as  
delivery_time,  
date_diff(order_delivered_customer_date,order_estimated_delivery_date,day) as  
diff_estimated_delivery  
FROM Abc_superstore.orders;
```

Sample output:

Row	order_id	order_delivered_customer_date	order_purchase_timestamp	delivery_time	diff_estimated_delivery
1	1950d777989f6a877539f5379...	2018-03-21 22:03:51 UTC	2018-02-19 19:48:52 UTC	30	12
2	2c45c33d2f9cb8ff8b1c86cc28...	2016-11-09 14:53:50 UTC	2016-10-09 15:39:56 UTC	30	-28
3	65d1e226dfaeb8cdc42f66542...	2016-11-08 10:58:34 UTC	2016-10-03 21:01:41 UTC	35	-16
4	635c894d068ac37e6e03dc54e...	2017-05-16 14:49:55 UTC	2017-04-15 15:37:38 UTC	30	-1
5	3b97562c3aee8bdedcb5c2e45...	2017-05-17 10:52:15 UTC	2017-04-14 22:21:54 UTC	32	0
6	68f47f50f04c4cb6774570cfde...	2017-05-16 09:07:47 UTC	2017-04-16 14:56:13 UTC	29	-1
7	276e9ec344d3bf029ff83a161c...	2017-05-22 14:11:31 UTC	2017-04-08 21:20:24 UTC	43	4
8	54e1a3c2b97fb0809da548a59...	2017-05-22 16:18:42 UTC	2017-04-11 19:49:45 UTC	40	4
9	fd04fa4105ee8045f6a0139ca5...	2017-05-19 13:44:52 UTC	2017-04-12 12:17:08 UTC	37	1
10	302bb8109d097a9fc6e9cefc5...	2017-05-23 14:19:48 UTC	2017-04-19 22:52:59 UTC	33	5
11	66057d37308e787052a32828...	2017-05-24 08:11:57 UTC	2017-04-15 19:22:06 UTC	38	6
12	19135c945c554eebfd7576c73...	2017-08-16 20:19:32 UTC	2017-07-11 14:09:37 UTC	36	2

5.2 Top 5 states with the highest & lowest average freight value.

Top 5 states with the **highest** average freight value:

Query:

```
SELECT s.seller_state, avg(freight_value) as top5_highest  
FROM Abc_superstore.order_items o  
JOIN Abc_superstore.sellers s ON o.seller_id = s.seller_id  
group by s.seller_state  
order by top5_highest desc  
limit 5;
```

Output:

Row	seller_state	top5_highest
1	RO	50.91285714285...
2	CE	46.38117021276...
3	PB	39.18815789473...
4	PI	36.94333333333...
5	AC	32.84

Insight :

From the above output, we can conclude that the states RO,CE,PB,PI & AC incurred the highest average freight charges.

Top 5 states with the **lowest** average freight value:

Query:

```
SELECT s.seller_state, avg(freight_value) as top5_lowest
FROM Abc_superstore.order_items o
JOIN Abc_superstore.sellers s ON
o.seller_id = s.seller_id
group by s.seller_state
order by top5_lowest
limit 5;
```

Output:

Row	seller_state	top5_lowest
1	SP	18.45221266585...
2	PA	19.38874999999...
3	RJ	19.47486508924...
4	DF	20.57181312569...
5	PR	22.72096874639...

Insight :

From the above output, we can conclude that the states SP, PA, RJ, DF, & PR incurred the lowest average freight charges.

5.3 The top 5 states with the highest & lowest average delivery time:

Top 5 states with the **highest** average delivery time:

Query :

```
SELECT c.customer_state,
ROUND(avg(date_diff(o.order_delivered_customer_date,o.order_purchase_timestamp,day)
),2) as highest_delivery_time
FROM Abc_superstore.orders o
JOIN Abc_superstore.customers c ON o.customer_id= c.customer_id
group by c.customer_state
order by highest_delivery_time desc
limit 5;
```

Output :

Row	customer_state	highest_delivery_time
1	RR	28.98
2	AP	26.73
3	AM	25.99
4	AL	24.04
5	PA	23.32

Insight:

The states RR, AP, AM, AL and PA have, on an average, taken the highest time to deliver the products.

- None of the above states appear in the top 12 states of customer distribution, as seen in Section 3.2. There may be a direct correlation between the delivery time and statewide customer distribution.

Top 5 states with the **lowest** average delivery time:

Query :

```
SELECT c.customer_state,
ROUND(avg(date_diff(o.order_delivered_customer_date,o.order_purchase_timestamp,day)
),2) as top5_lowest
FROM Abc_superstore.orders o
JOIN Abc_superstore.customers c ON o.customer_id= c.customer_id
group by c.customer_state
order by top5_lowest
limit 5;
```

Output :

Row	customer_state	lowest_delivery_time
1	SP	8.3
2	PR	11.53
3	MG	11.54
4	DF	12.51
5	SC	14.48

Insight :

The states SP, PR, MG, DF, and SC have, on an average, taken the lowest time to deliver the products.

5.4 The top 5 states where the order delivery is really fast as compared to the estimated date of delivery.

Query:

```
SELECT c.customer_state,
ROUND(AVG(date_diff(order_estimated_delivery_date,order_delivered_customer_date,day
)),2) as fast_states
FROM Abc_superstore.orders o
JOIN Abc_superstore.customers c ON o.customer_id= c.customer_id
GROUP BY c.customer_state
ORDER BY fast_states desc
LIMIT 5;
```

Output:

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

Insight :

The states AC, RO, AP, AM, and RR have the fastest average delivery times.

6. Analysis based on the payments:

6.1 The month on month no. of orders placed using different payment types.

Query :

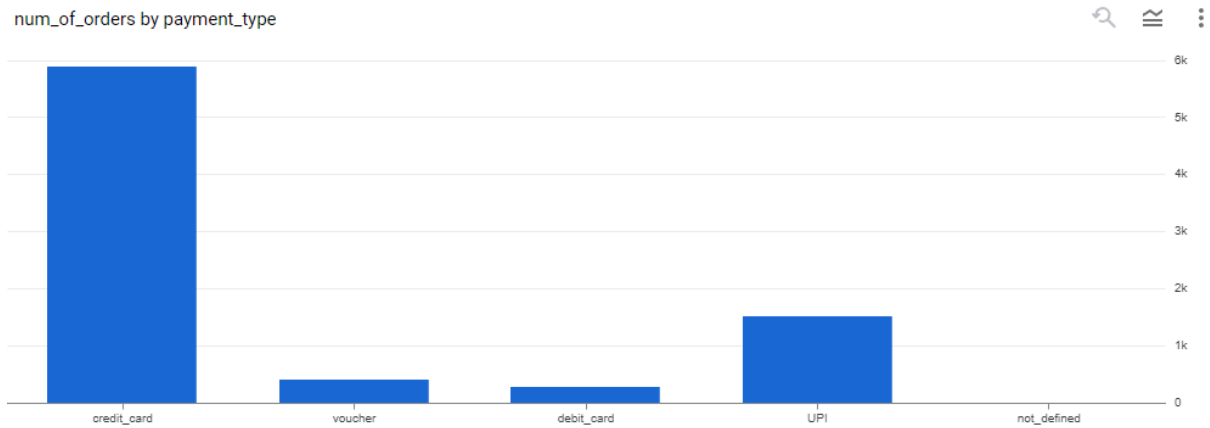
```
WITH cte as
(
SELECT payment_type,
EXTRACT(month from order_purchase_timestamp) as month,
EXTRACT(year from order_purchase_timestamp) as year,
count(*) as num_of_orders
FROM Abc_superstore.payments p
JOIN Abc_superstore.orders o USING(order_id)
GROUP BY payment_type,
EXTRACT(month from order_purchase_timestamp),
EXTRACT(year from order_purchase_timestamp)
)
SELECT * FROM cte
ORDER BY cte.year,cte.month;
```

Output :

Row	payment_type	month	year	num_of_orders
1	credit_card	9	2016	3
2	credit_card	10	2016	254
3	voucher	10	2016	23
4	debit_card	10	2016	2
5	UPI	10	2016	63
6	credit_card	12	2016	1
7	voucher	1	2017	61
8	UPI	1	2017	197
9	credit_card	1	2017	583
10	debit_card	1	2017	9
11	credit_card	2	2017	1356
12	voucher	2	2017	119

Insight :

We can see that payments have been made using different means such as credit card, debit card, voucher, and UPI.



From the above graph, it can be concluded that most of the payments have been made through credit cards, followed by UPI, vouchers and debit cards.

6.2 The no. of orders placed on the basis of the payment installments that have been paid:

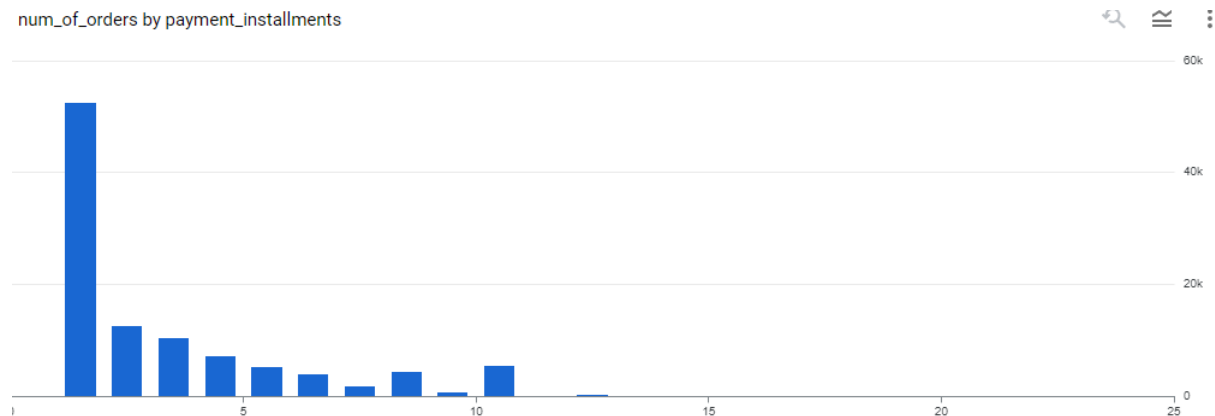
Query :

```
SELECT payment_installments, count(order_id) as num_of_orders
FROM Abc_superstore.payments
GROUP BY payment_installments HAVING payment_installments >= 1
ORDER BY payment_installments;
```

Output :

Row	payment_installment	num_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

Insight :



We can see that most of the payments have been made through a single installment, and the number of orders made decrease as the installments increase.

Recommendation:

Low cost EMIs can be provided for longer time durations.

Summary : Insights and Recommendations

Section 3.2 reveals that the state SP has the highest number of customers, with almost 3 times more than the second highest state – RJ, which is followed closely by the state MG. The 4th highest state, RS, has lesser than half the number of customers from MG.

- There is still a huge scope for growth in many other states.

Section 4.2 shows that the states PB, AL, PA, AC have high average order price values. It may indicate that the states have more affluent customers. For example, the state with the 12th highest total order price, CE, has an average order price of 153.76 despite not making it to the top 10 states in total value of order price.

- More luxury based items can be sold in these states where the customers have higher buying power.

Section 5.3 shows that the states RR, AP, AM, AL and PA have, on an average, taken the highest time to deliver the products. The average time taken for delivery in these states ranges from 29 days to 23 days.

- None of the above states appear in the top 12 states of customer distribution, as seen in Section 3.2. There may be a **direct correlation** between the **delivery time** and **statewise customer distribution**.
- The delivery time must be improved in these states.

Section 6.2 indicates that most of the payments have been made through a single installment, and the number of orders made decrease as the number installments increase.

- Low cost EMI can be provided for longer time durations.

Abc_superstore is a favourite brand amongst thousands of customers across Brazil. It has shown a significant rise in revenue over the years, as seen in **Section 2.1**. The customers loyalty can also be verified by the fact that there is an increase of 138.53% in the cost of orders placed between Jan-Aug, 2017 & 2018, as seen in **Section 4.1**.

- Marketing can be done in the states with low customers, based on the customer satisfaction of the thousands of happy customers from the states with highest customers.