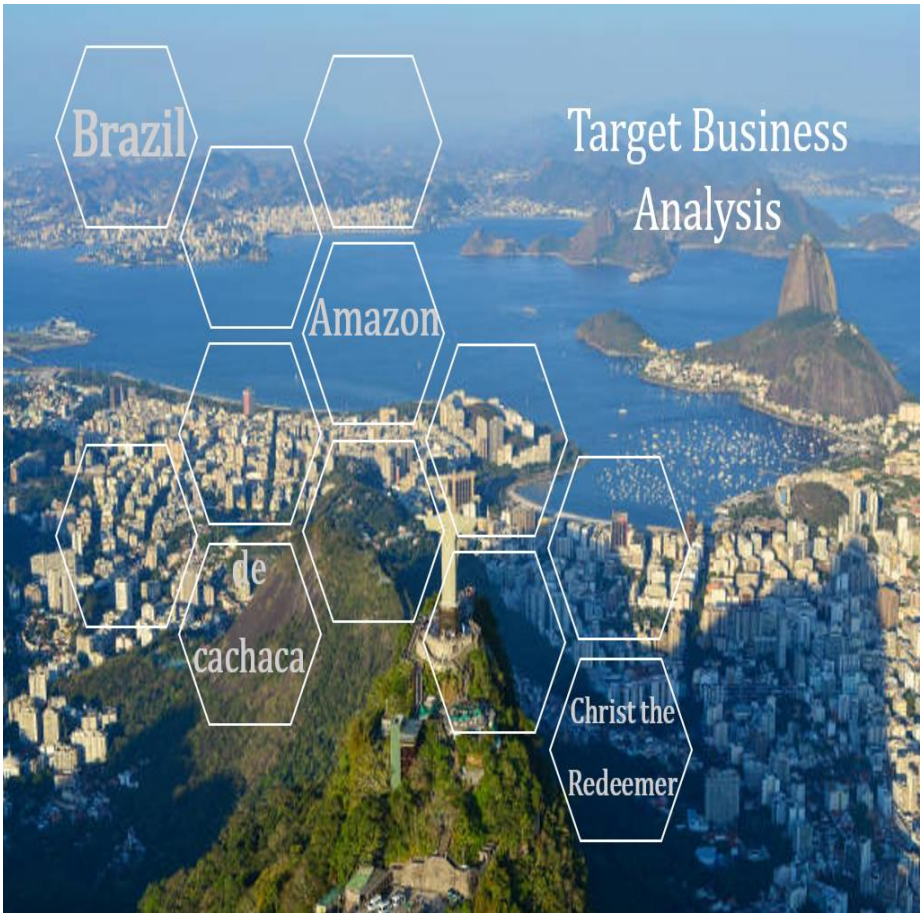


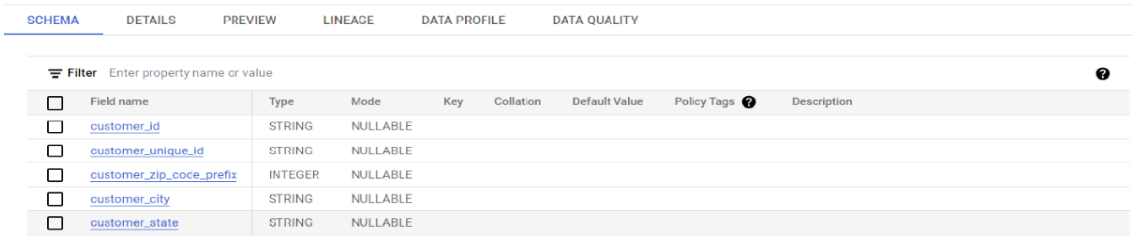
BUISNESS ANALYSIS CASE STUDY

ABSTRACT
The purpose of the paper is to present a structural analysis of market research of TARGET .

Manjunatha B B

1. Usual Exploratory Data Analysis

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

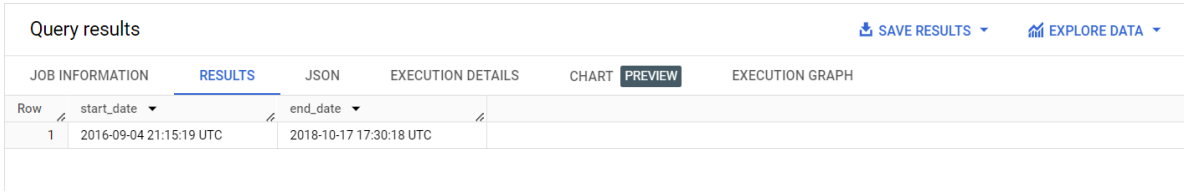


Field name	Type	Mode	Key	Collation	Default Value	Policy Tags	Description
customer_id	STRING	NULLABLE					
customer_unique_id	STRING	NULLABLE					
customer_zip_coce_prefix	INTEGER	NULLABLE					
customer_city	STRING	NULLABLE					
customer_state	STRING	NULLABLE					

- By looking into the schema of "customers" table it is understood that, the customer_id, customer_unique_id, customer_city and customer_state are of string data type and null values are allowed.
- Where in customer_zip_code_prefix is integer data type and nullable

1.2 Get the time range between which the orders were placed.

```
select min(order_purchase_timestamp) start_date,
    max(order_purchase_timestamp) end_date
from `brazilian_market.orders` ;
```

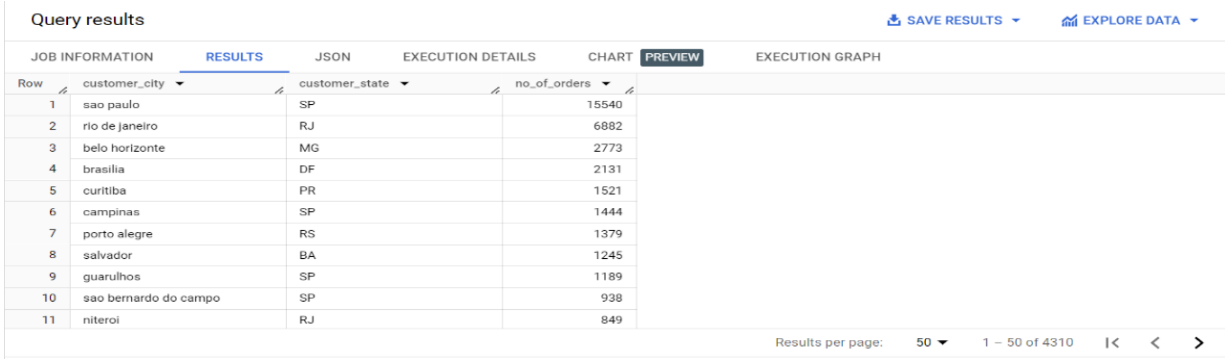


Row	start_date	end_date
1	2016-09-04 21:15:19 UTC	2018-10-17 17:30:18 UTC

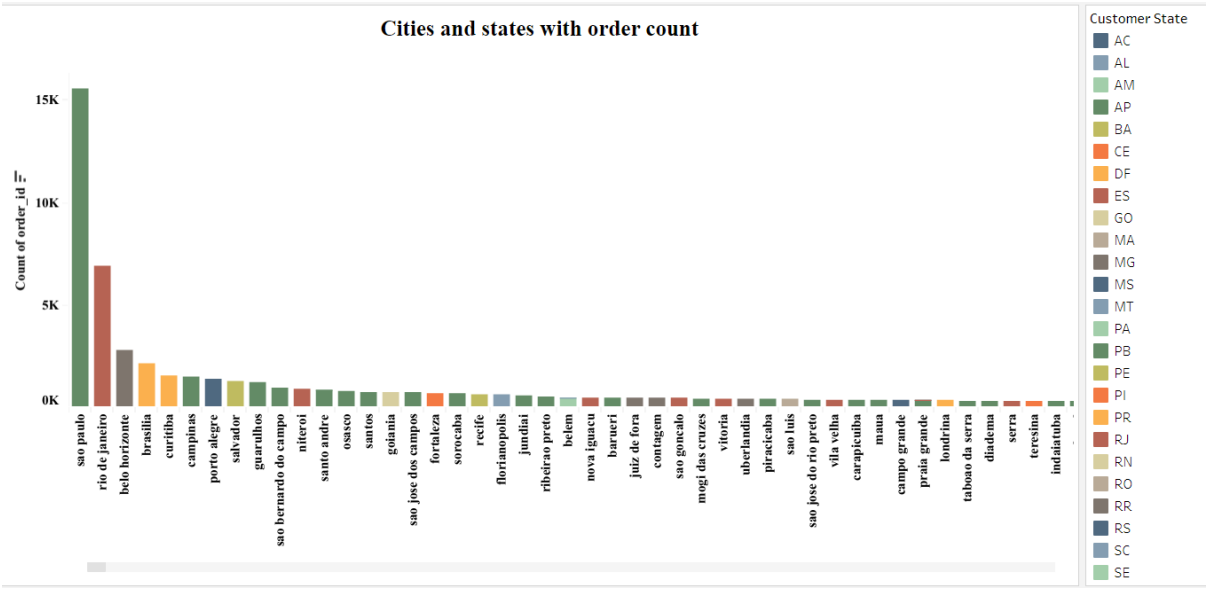
- The dataset that we are analysing in the range of dates between 04th September 2016 and 17th October 2018

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

```
select distinct C.customer_city, C.customer_state,
    count(O.customer_id) no_of_orders
from `brazilian_market.orders` O
left join `brazilian_market.customers` C
on O.customer_id = C.customer_id
group by 1,2
order by 3 desc ;
```



Row	customer_city	customer_state	no_of_orders
1	sao paulo	SP	15540
2	rio de janeiro	RJ	6882
3	belo horizonte	MG	2773
4	brasilia	DF	2131
5	curitiba	PR	1521
6	campinas	SP	1444
7	porto alegre	RS	1379
8	salvador	BA	1245
9	guarulhos	SP	1189
10	sao bernardo do campo	SP	938
11	niteroi	RJ	849



- Here, Sao Paulo city from SP state alone itself have orders more than the following five cites combined. This is because of Sao Paulo is the most populated and richest state in Brazil.

2. In-depth Exploration

2.1 Is there a growing trend in the no. of orders placed over the past years?

```
select extract(year from order_purchase_timestamp) year,
    extract (month from order_purchase_timestamp) month,
    count(distinct order_id) order_count
from `brazilian_market.orders`
group by 1,2
order by 1, 2 ;
```

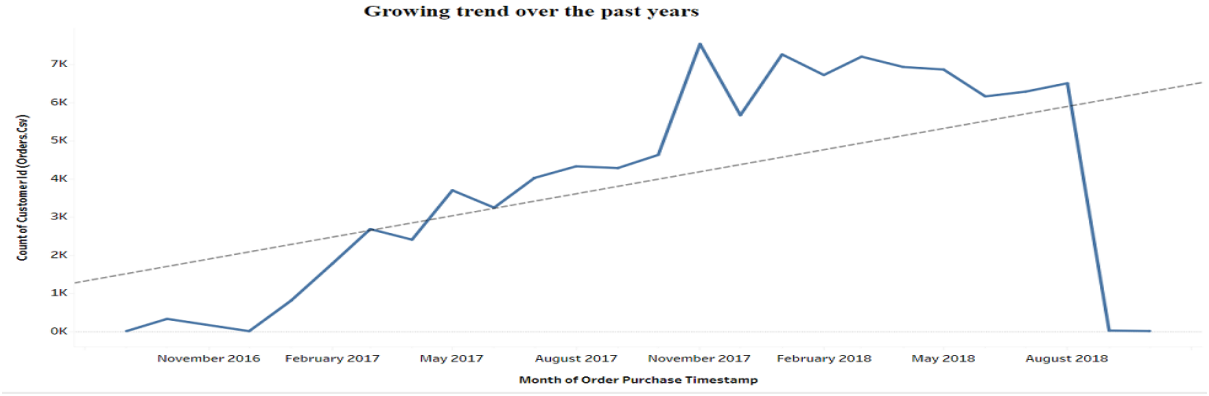
Query results

SAVE RESULTS | EXPLORE DATA

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS | CHART PREVIEW | EXECUTION GRAPH

Row	year	month	order_count
1	2016	9	4
2	2016	10	324
3	2016	12	1
4	2017	1	800
5	2017	2	1780
6	2017	3	2682
7	2017	4	2404
8	2017	5	3700
9	2017	6	3245
10	2017	7	4026
11	2017	8	4331

Results per page: 50 | 1 – 25 of 25



- By analysing count of orders, we observed that there is a growing trend in Brazil E-commerce market. The count of orders shows overall upward trend with some fluctuations

2.2 Can we see some kind of monthly seasonality in terms of the no. of orders being placed?

```
select extract(month from order_purchase_timestamp) month,
    count(distinct order_id) order_count
from `brazilian_market.orders`
group by month
order by month ;
```

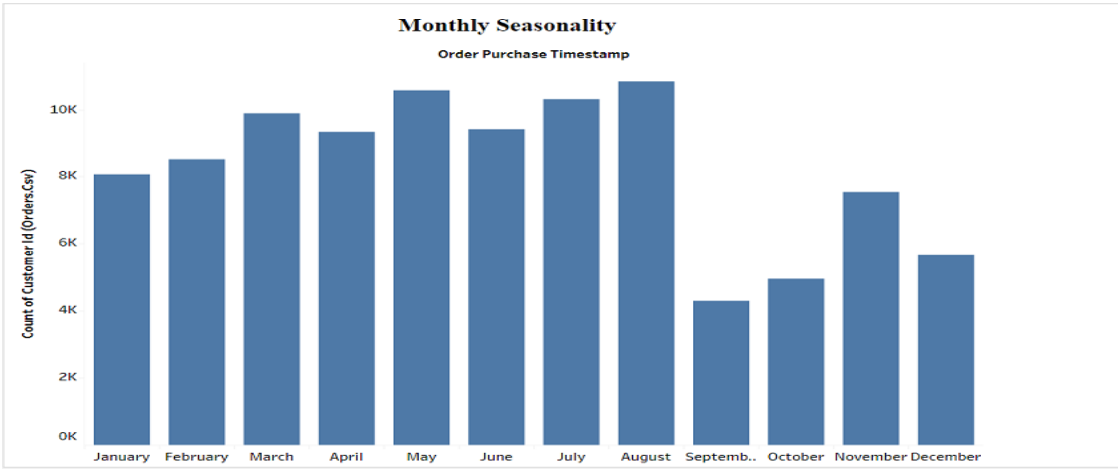
Query results

SAVE RESULTS | EXPLORE DATA

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS | CHART PREVIEW | EXECUTION GRAPH

Row	month	order_count
1	1	8069
2	2	8508
3	3	9893
4	4	9343
5	5	10573
6	6	9412
7	7	10318
8	8	10843
9	9	4305
10	10	4959
11	11	7544

Results per page: 50 1 – 12 of 12

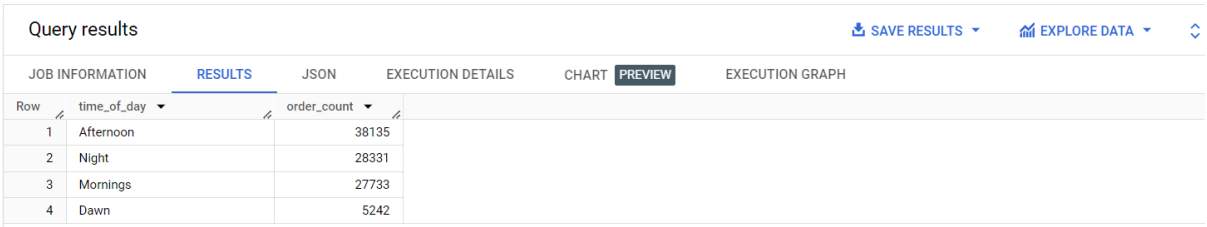


- By analysing and visualizing the graph, we can understand that there is some seasonality in E-commerce in Brazil over the months.
- The count of orders generally increases from March to August with some fluctuations, But the highest number of orders(peak) recorded in August month its because of the festival de cachaca dedicated to national liquor.
- It is also noted that there is an increase in order count in February and March, it's because of Carnival season in Brazil.

2.3 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

```
select
    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'
      else 'Night' end time_of_day,
    count(order_id) order_count
from `brazilian_market.orders`
group by 1
order by 2 desc ;
```



Query results

SAVE RESULTS | EXPLORE DATA

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS | CHART PREVIEW | EXECUTION GRAPH

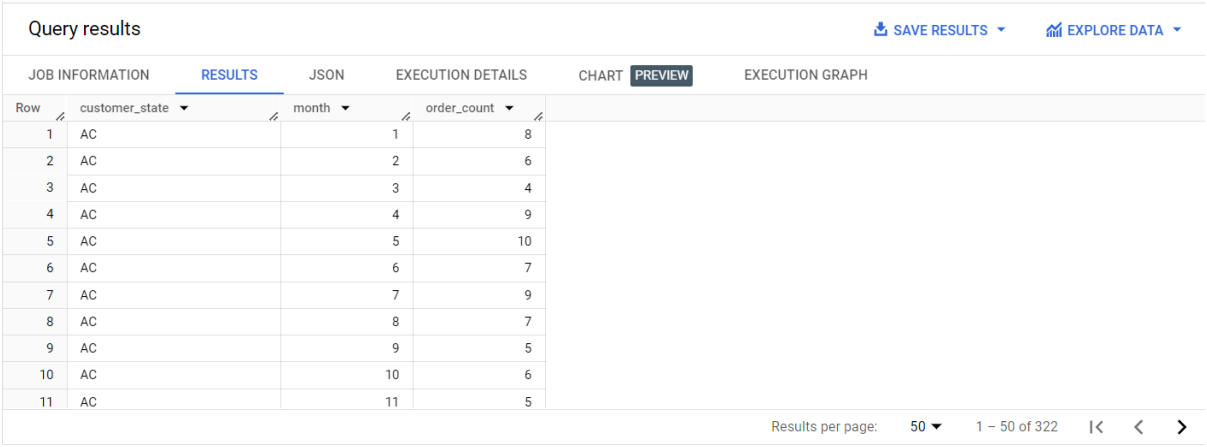
Row	time_of_day	order_count
1	Afternoon	38135
2	Night	28331
3	Mornings	27733
4	Dawn	5242

- We found that Brazilian customers tends to order more on Day time, special in afternoon and night.
- This shows that customers prefer to shop when they have leisure time or after finishing there daily activities.

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

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

```
select  C.customer_state,
      extract(month from O.order_purchase_timestamp) month,
      count(O.order_id) order_count
from `brazilian_market.orders` O
left join `brazilian_market.customers` C
on O.customer_id = C.customer_id
group by 1, 2
order by 1, 2 ;
```



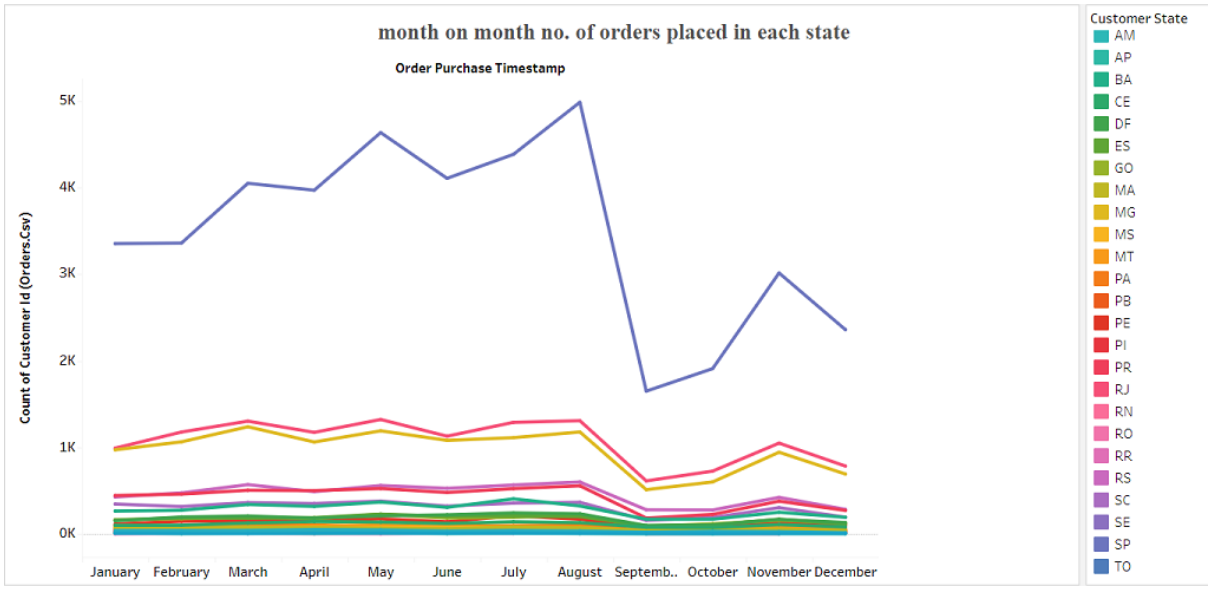
Query results

SAVE RESULTS | EXPLORE DATA

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS | CHART PREVIEW | EXECUTION GRAPH

Row	customer_state	month	order_count
1	AC	1	8
2	AC	2	6
3	AC	3	4
4	AC	4	9
5	AC	5	10
6	AC	6	7
7	AC	7	9
8	AC	8	7
9	AC	9	5
10	AC	10	6
11	AC	11	5

Results per page: 50 | 1 – 50 of 322



- By illustrating the month-on-month number of orders placed in different states in Brazil, it is evident that Sao Paulo (SP) has the highest orders in every month followed by Rio de Janeiro (RJ) and Minas Gerais (MG).

3.2 How are the customers distributed across all the states?

```
select customer_state ,
    count(distinct customer_id) customer_count
from `brazilian_market.customers`
group by 1
order by 2 desc ;
```

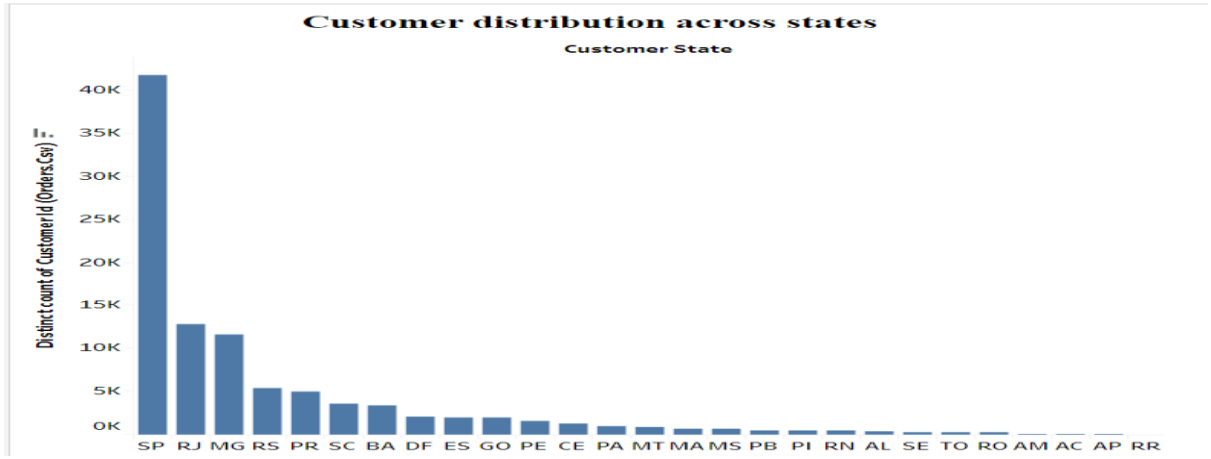
Query results

SAVE RESULTS | EXPLORE DATA

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS | CHART PREVIEW | EXECUTION GRAPH

Row	customer_state	customer_count
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

Results per page: 50 | 1 – 27 of 27



- The data revels that Sao Paulo state has the highest number of customers since it is a most populated state in the country.

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

4.1 Get the % increase in the cost of orders from year 2017 to 2018 (include months between Jan to Aug only)

Percentage increase over the months

```
with cte as
(select year, month , sum(payment_value) Tot_sum
from
(select O.order_id, P.payment_value, O.order_purchase_timestamp,
      extract(month from  O.order_purchase_timestamp) month,
      extract(year from  O.order_purchase_timestamp) year
from `brazilian_market.orders` O
join `brazilian_market.payments` P
on O.order_id = P.order_id
where extract(month from  O.order_purchase_timestamp) between 1 and 8)A
group by 1,2
order by 1,2 ),

cte1 as
(select *,
     lead(Tot_sum,8) over(order by year,month) lead_sum
from cte
order by year,month)

select month,
    round(((lead_sum - Tot_sum)/Tot_sum) * 100,2) as percent_increase
from cte1
limit 8 ;
```

Query results

SAVE RESULTS

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS | CHART PREVIEW | EXECUTION GRAPH

Row	month	percent_increase
1	1	705.13
2	2	239.99
3	3	157.78
4	4	177.84
5	5	94.63
6	6	100.26
7	7	80.04
8	8	51.61

Percentage increase over the year

```
with cte as
(select year,  sum(payment_value) as total_value
from
(select P.order_id, P.payment_value,
     O.order_purchase_timestamp, extract(year from O.order_purchase_timestamp) as year
from `brazilian_market.payments` P
```

```
join `brazilian_market.orders` O
on P.order_id = O.order_id
where O.order_purchase_timestamp between '2017-01-01' and '2017-08-31'
  or  O.order_purchase_timestamp between '2018-01-01' and '2018-08-31')A
group by year ),

appended_value as
(select *, lead(total_value) over( order by year) app_total_value
from cte )

select year ,
round(((app_total_value - total_value)/ total_value) * 100,2) as perc_increase
from appended_value ;
```

Query results

SAVE RESULTS

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS | CHART PREVIEW | EXECUTION GRAPH

Row	year	perc_increase
1	2018	null
2	2017	138.53

- The overall percentage increase in the cost of orders from 2017 to 2018, only including the months from January to August, is 138.53%. where in month wise January shows the highest percentage increase followed by February and April.

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

```
select customer_state, round(sum(price),2) Toatal_price,
    round(avg(price),2) Average_price
from
(select O.order_id, I.price, I.freight_value ,C.customer_state,
from `brazilian_market.orders` O
join `brazilian_market.order_items` I
on O.order_id = I.order_id
join `brazilian_market.customers` C
on O.customer_id = C.customer_id)A
group by 1 ;
```

Query results

SAVE RESULTS | EXPLORE DATA

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS | CHART PREVIEW | EXECUTION GRAPH

Row	customer_state	Toatal_price	Average_price
1	MT	156453.53	148.3
2	MA	119648.22	145.2
3	AL	80314.81	180.89
4	SP	5202955.05	109.65
5	MG	1585308.03	120.75
6	PE	262788.03	145.51
7	RJ	1824092.67	125.12
8	DF	302603.94	125.77
9	RS	750304.02	120.34
10	SE	58920.85	153.04
11	PR	683083.76	119.0

Results per page: 50 | 1 – 27 of 27

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

```
select customer_state, round(sum(freight_value),2) Toatal_freight_value,
    round(avg(freight_value),2) Average_freight_value
from
(select O.order_id, I.price, I.freight_value ,C.customer_state,
from `brazilian_market.orders` O
join `brazilian_market.order_items` I
on O.order_id = I.order_id
join `brazilian_market.customers` C
on O.customer_id = C.customer_id)A
group by 1
```

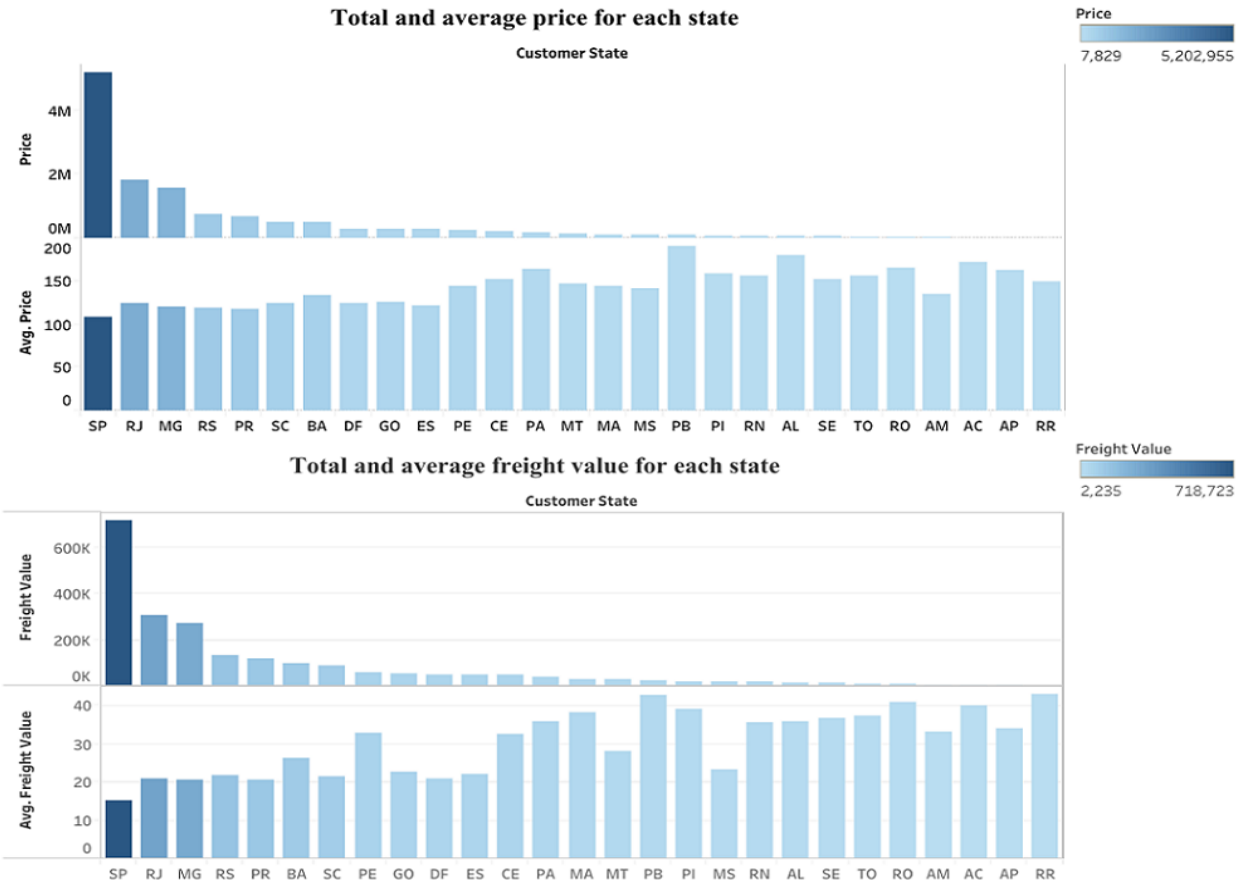
Query results

SAVE RESULTS | EXPLORE DATA

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS | CHART PREVIEW | EXECUTION GRAPH

Row	customer_state	Toatal_freight_value	Average_freight_valu
1	MT	29715.43	28.17
2	MA	31523.77	38.26
3	AL	15914.59	35.84
4	SP	718723.07	15.15
5	MG	270853.46	20.63
6	PE	59449.66	32.92
7	RJ	305589.31	20.96
8	DF	50625.5	21.04
9	RS	135522.74	21.74
10	SE	14111.47	36.65
11	PR	117851.68	20.53

Results per page: 50 1 – 27 of 27

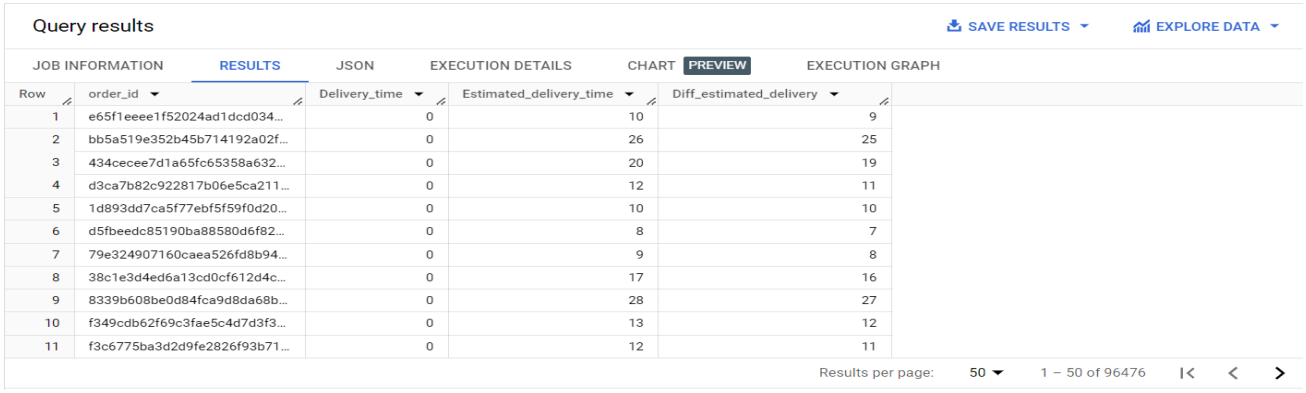


- By analysing and visualizing the graphs it is noticed that, while Sao Paulo has the highest total price value and total freight value, But surprisingly lowest average price value and average freight value.
- Paraiba (PB) has the highest average price value and average freight value.

5. Analysis based on sales, freight and delivery time

5.1 Find the no. of days taken to deliver each order from the order's purchase date as delivery time

```
select order_id,
    date_diff(order_delivered_customer_date, order_purchase_timestamp, day) Delivery_time,
    date_diff(order_estimated_delivery_date, order_purchase_timestamp, day) Estimated_delivery_time,
    date_diff(order_estimated_delivery_date, order_delivered_customer_date, day) Diff_estimated_delivery
from `brazilian_market.orders`
where date_diff(order_delivered_customer_date, order_purchase_timestamp, day) is not null
order by 2 ;
```



Query results

SAVE RESULTS | EXPLORE DATA

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS | CHART PREVIEW | EXECUTION GRAPH

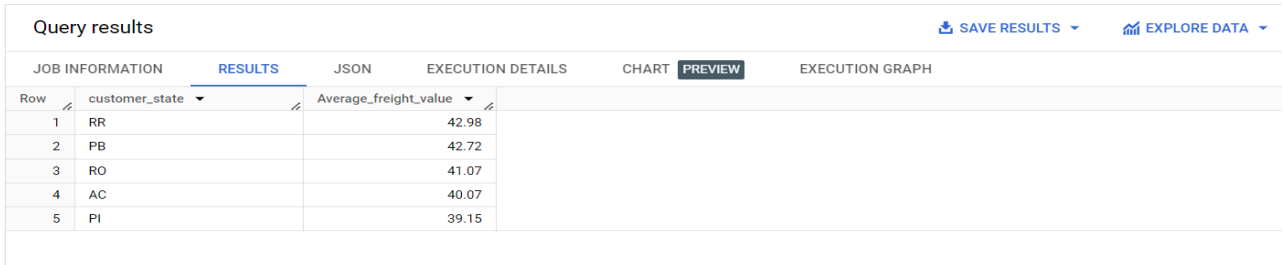
Row	order_id	Delivery_time	Estimated_delivery_time	Diff_estimated_delivery
1	e65f1eeee1f52024ad1dcd034...	0	10	9
2	bb5a519e352b45b714192a02f...	0	26	25
3	434cecee7d1a65fc65358a632...	0	20	19
4	d3ca7b82c922817b06e5ca211...	0	12	11
5	1d893dd7ca5f77ebf5f59f0d20...	0	10	10
6	d5fbeedc85190ba88580d6f82...	0	8	7
7	79e324907160caea526fd8b94...	0	9	8
8	38c1e3d4ed6a13cd0cf612d4c...	0	17	16
9	8339b608be0d84fca9d8da68b...	0	28	27
10	f349cdb62f69c3fae5c4d7d3f3...	0	13	12
11	f3c6775ba3d2d9fe2826f93b71...	0	12	11

Results per page: 50 | 1 – 50 of 96476

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

Top 5 highest average freight value

```
select customer_state,
    round(avg(freight_value),2) Average_freight_value
from
(select O.order_id, I.price, I.freight_value ,C.customer_state,
from `brazilian_market.orders` O
join `brazilian_market.order_items` I
on O.order_id = I.order_id
join `brazilian_market.customers` C
on O.customer_id = C.customer_id)A
group by 1
order by 2 desc
limit 5 ;
```



Query results

SAVE RESULTS | EXPLORE DATA

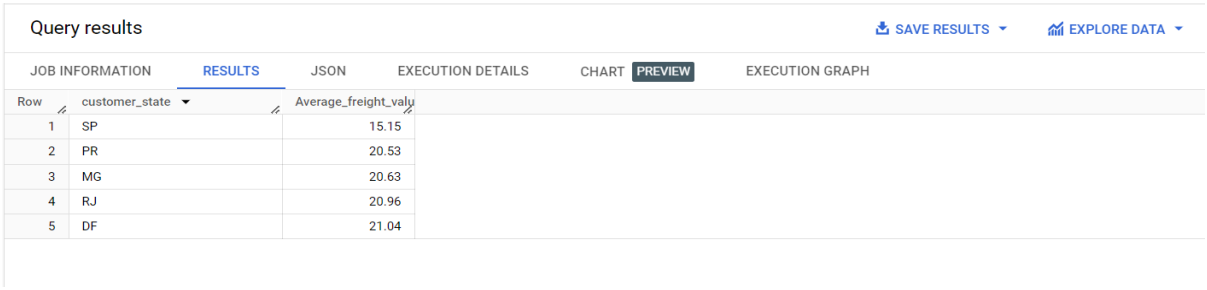
JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS | CHART PREVIEW | EXECUTION GRAPH

Row	customer_state	Average_freight_value
1	RR	42.98
2	PB	42.72
3	RO	41.07
4	AC	40.07
5	PI	39.15

- Roraima (RR) and Paraiba (PB) has the highest average freight value.

Top 5 lowest average freight value

```
select customer_state,
    round(avg(freight_value),2) Average_freight_value
from
(select O.order_id, I.price, I.freight_value ,C.customer_state,
from `brazilian_market.orders` O
join `brazilian_market.order_items` I
on O.order_id = I.order_id
join `brazilian_market.customers` C
on O.customer_id = C.customer_id)A
group by 1
order by 2
limit 5 ;
```



Query results

SAVE RESULTS | EXPLORE DATA

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS | CHART PREVIEW | EXECUTION GRAPH

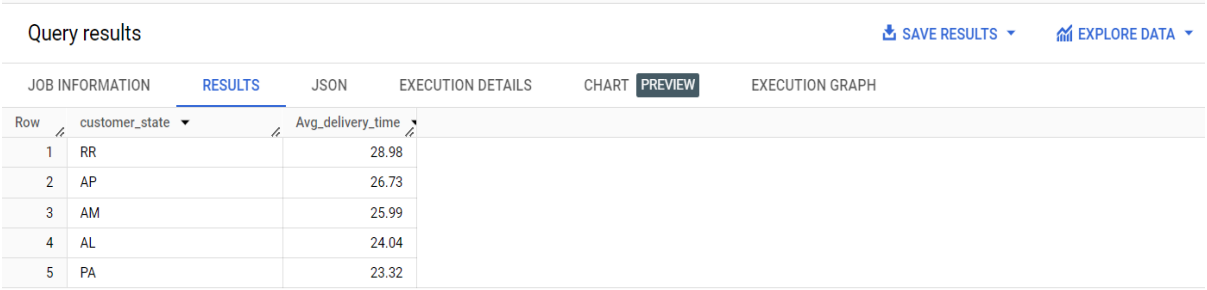
Row	customer_state	Average_freight_valu
1	SP	15.15
2	PR	20.53
3	MG	20.63
4	RJ	20.96
5	DF	21.04

- Sao Paulo (SP) has the lowest average freight value.

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

Top 5 states with the highest average delivery time

```
select C.customer_state,
    round(avg(date_diff(O.order_delivered_customer_date, O.order_purchase_timestamp, day)),2)
Avg_delivery_time
from `brazilian_market.orders` O
join `brazilian_market.customers` C
on O.customer_id = C.customer_id
where date_diff(O.order_delivered_customer_date, O.order_purchase_timestamp, day) is not null
group by 1
order by 2 desc
limit 5 ;
```



Query results

SAVE RESULTS | EXPLORE DATA

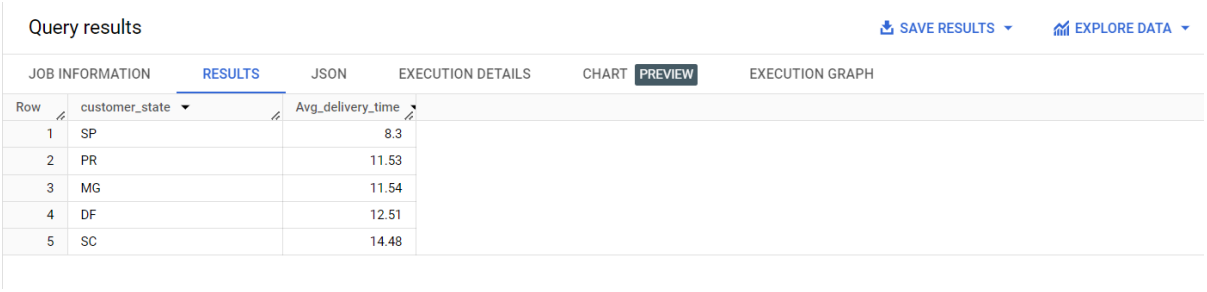
JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS | CHART PREVIEW | EXECUTION GRAPH

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

- Roraima (RR) has the highest average delivery time. For an improvement of an business delivery time should be as less as possible.

Top 5 states with the lowest average delivery time

```
select C.customer_state,
    round(avg(date_diff(O.order_delivered_customer_date, O.order_purchase_timestamp, day)),2)
Avg_delivery_time
from `brazilian_market.orders` O
join `brazilian_market.customers` C
on O.customer_id = C.customer_id
where date_diff(O.order_delivered_customer_date, O.order_purchase_timestamp, day) is not null
group by 1
order by 2
limit 5 ;
```

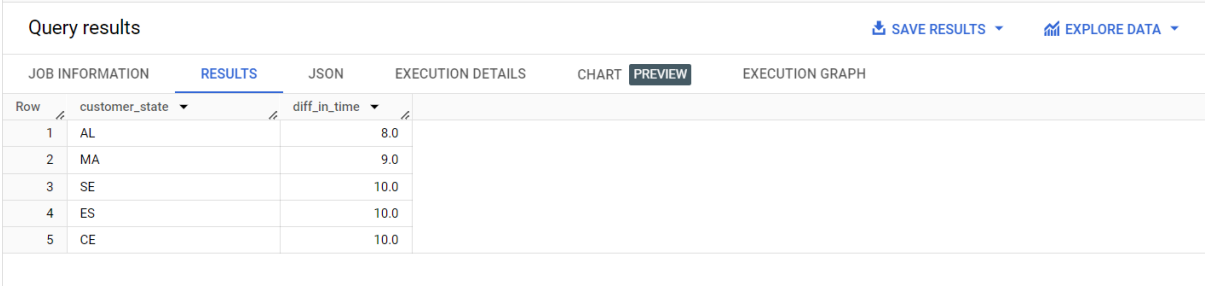


Row	customer_state	Avg_delivery_time
1	SP	8.3
2	PR	11.53
3	MG	11.54
4	DF	12.51
5	SC	14.48

- Sao Paulo (SP) has the lowest average delivery time. This is the good sign of improvement of the business.

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

```
select C.customer_state,
    ceil(avg(date_diff(O.order_estimated_delivery_date, O.order_delivered_customer_date, day)))
diff_in_time
from `brazilian_market.orders` O
join `brazilian_market.customers` C
on O.customer_id = C.customer_id
where date_diff(O.order_delivered_customer_date, O.order_estimated_delivery_date, day) is not null
group by 1
order by 2
limit 5 ;
```



Row	customer_state	diff_in_time
1	AL	8.0
2	MA	9.0
3	SE	10.0
4	ES	10.0
5	CE	10.0

- Alagoas (AL) has order delivered really fast as compared to the estimate date of delivery, followed by Maranhao (MA) and Sergipe (SE).
- This indicates that other than Sao Paulo (SP), another states in Brazil also on the way of improving the business.

6. Analysis based on the payments

6.1 Find the month-on-month no. of orders placed using different payment types.

```
select P.payment_type,
    extract(month from order_purchase_timestamp) as month,
    count(payment_type) No_of_orders
from `brazilian_market.payments`  P
join `brazilian_market.orders` O
on P.order_id = O.order_id
where payment_type  != 'not_defined'
group by 1, 2
order by 1,2 ;
```

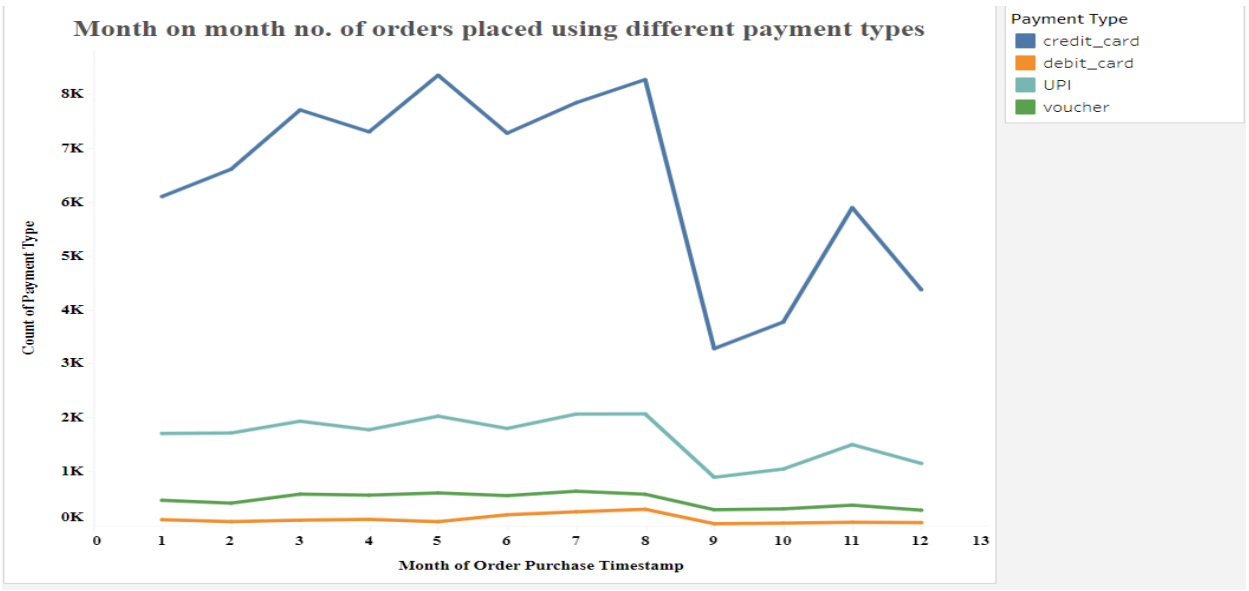
Query results

SAVE RESULTS | EXPLORE DATA

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS | CHART PREVIEW | EXECUTION GRAPH

Row	payment_type	month	No_of_orders
1	UPI	1	1715
2	UPI	2	1723
3	UPI	3	1942
4	UPI	4	1783
5	UPI	5	2035
6	UPI	6	1807
7	UPI	7	2074
8	UPI	8	2077
9	UPI	9	903
10	UPI	10	1056
11	UPI	11	1509

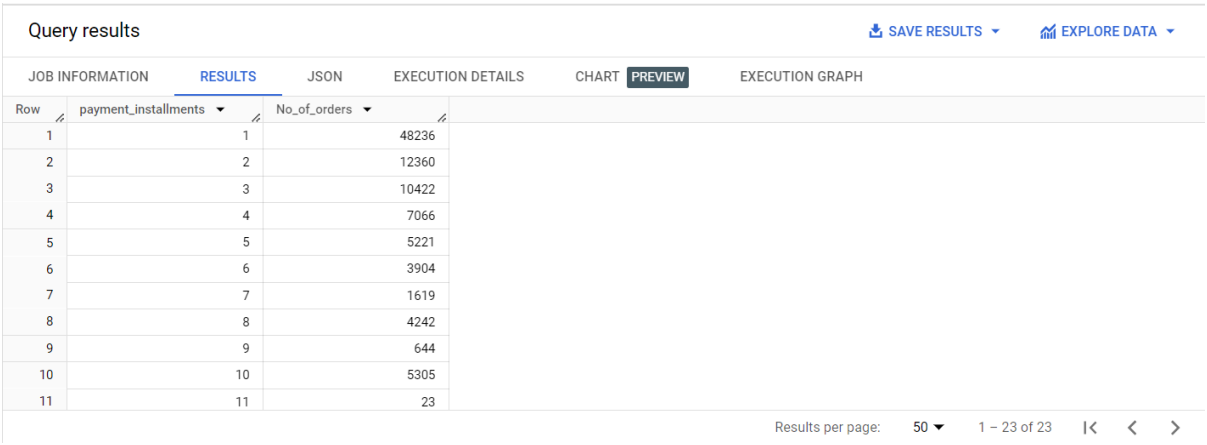
Results per page: 50 | 1 – 48 of 48



- The analysis shows that overall uptrend from January to August and another uptrend from September to November.
- Credit card transactions are the most preferred way of payment method followed by UPI.
- Where in debit cards are the least preferred way of payment method.

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

```
select P.payment_installments,
    count(P.payment_installments) No_of_orders
from `brazilian_market.payments` P
join `brazilian_market.orders` O
on P.order_id = O.order_id
where  P.payment_sequential = 1
group by 1 ;
```



Row	payment_installments	No_of_orders
1	1	48236
2	2	12360
3	3	10422
4	4	7066
5	5	5221
6	6	3904
7	7	1619
8	8	4242
9	9	644
10	10	5305
11	11	23

- This is revels that majority of the orders have one payment installation (48236 maximum count).
- The highest number of installments is 24, which is associated with 18 orders.

RECOMMENDATIONS

- The state SP has significantly high number of orders than next five states combined. This indicates that there is an opportunity for improvements in other states. Focusing on these can help to improve the number of orders and also can increase the customer base.

- Monthly variations in sales are observed with high peaks in festival periods. If a business can focus on marketing and sales strategies according to enhance the business capital at these peak periods will improve the sales growth.

- Reducing the delivery time in areas with the longer delivery duration will have positive impact on customer satisfaction.

- The data indicates that decline in orders during September and October. Offering some discounts in off peak seasons will boost the sales.

- Improve logistics and shipping process to reduce delivery time and enhance the customer satisfaction.

- Evaluate pricing and freight fees to ensure competitiveness in the market while maximizing revenue and profitability.

- Invest in technology and infrastructure to enhance the E-Commerce experience. Like improving website performance and providing some personalized product recommendations based on customer purchase behaviour.

- Utilize social media platforms to improve the brand awareness.

- States like Sao Paulo (SP) and Rio de Janeiro (RJ) have high order count to further boost the sales focus on customer retention strategies like personalized marketing campaigns.

- Monitor compotator activity and adjust the business strategies accordingly, such as offering better pricing, expanding product offering or improve customer service to stay competitive in market.

By implementing these recommendations, businesses can optimize their operations, enhance customer satisfaction, and drive overall sales growth in the Brazilian e-commerce market.