Business Case: Target SQL

(Raviteja Dommeti) (ravitejadommeti092@gmail.com)

1.Import the dataset and do usual exploratory analysis steps like checking the structure & characteristics of the dataset

1)1) Data type of columns in a table::

Data Set Name	Feature	Dtype
Customer	Customer_id	object
	Customer_unique_id	object
	customer_zip_code_prefix	int64
	customer_city	object
	customer_state	object
Geolocation	geolocation_zip_code_prefix	int64
	geolocation_lat	float64
	geolocation_lng	float64
	geolocation_city	object
	geolocation_state	object
Orders	order_delivered_customer_date	object
	order_delivered_carrier_date	object
	order_approved_at	object
	order_id	object
	customer_id	object
	order_status	object
	order_purchase_timestamp	object
	order_estimated_delivery_date	object
Order_items	order_id	object
	order_item_id	int64
	product_id	object
	seller_id	object
	shipping_limit_date	object

	price	float64
	freight_value	float64
Order_payments	order_id	object
	payment_sequential	int64
	payment_type	object
	payment_installments	int64
	payment_value	float64
Order_reviews	review_comment_title	object
	review_comment_message	object
	review_id	object
	order_id	object
	review_score	int64
	review_creation_date	object
	review_answer_timestamp	object
Products	product_category_name	object
	product_name_lenght	float64
	product_description_lenght	float64
	product_photos_qty	float64
	product_weight_g	float64
	product_length_cm	float64
	product_height_cm	float64
	product_width_cm	float64
	product_id	object
Sellers	seller_id	object
	seller_zip_code_prefix	int64
	seller_city	object
	seller_state	object

1)2) Time period for which the data is given::

```
select min(order_purchase_timestamp) as min_date, max(order_purchase_timestamp) as
max_date from `sql-project-380605.Company.orders`
```

Efficient Time Period is in between 2016-09 and 2018-10.

1)3) Cities and States of customers ordered during the given period::

```sql
SELECT distinct customer_city as city ,customer_state as state  FROM
`sql-project-380605.Company.customers` order by 2,1
```


So,From the data We can conclude that during the given Period the Company had Received orders from around 4310 cities from 27 States .

2)IN DEPTH EXPLORATION :

2)2) What time do Brazilian customers tend to buy (Dawn, Morning, Afternoon or Night)?

```sql
select count(order_id) as total_orders_by_time_day,time_day  from
(select order_id ,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"Morning"
when  extract(hour from order_purchase_timestamp) between 13 and 18 then"Afternoon"
else "Night" end as time_day from `sql-project-380605.Company.orders` ) as tbl
group by 2 order by 1 desc;
```
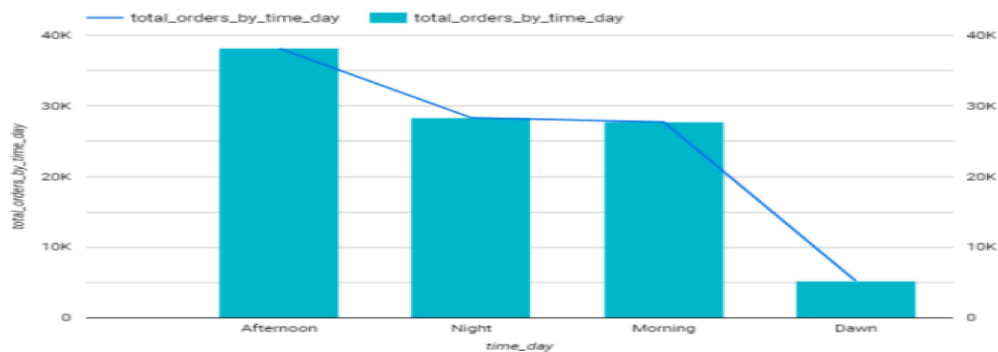
- From this we can conclude that brazilian customers tend to buy more in afternoon followed by night,morning,dawn

2)1)Is there a growing trend of e-commerce in Brazil? How can we describe a complete scenario? Can we see some seasonality with peaks at specific months?

```sql
select year,month ,round(sum(payment_value) , 2) as total_amount ,
round(count(distinct order_id), 2) as total_orders from
(SELECT o.*,p.payment_value,extract(year from o.order_purchase_timestamp) as year,
extract(month from o.order_purchase_timestamp) as month FROM
`sql-project-380605.Company.orders` as o  join
`sql-project-380605.Company.payments` as p on o.order_id=p.order_id
order by p.payment_value desc) as tbl group by 1,2 order by 4 desc
```


Row	year	month	total_amount	total_orders
1	2017	11	1194882.8	7544.0
2	2018	1	1115004.18	7269.0
3	2018	3	1159652.12	7211.0
4	2018	4	1160785.48	6939.0
5	2018	5	1153982.15	6873.0
6	2018	2	992463.34	6728.0
7	2018	8	1022425.32	6512.0
8	2018	7	1066540.75	6292.0
9	2018	6	1023880.5	6167.0

Evolution of Total Orders in Brazilian E-Commerce

E-commerce in Brazil has been on the rise throughout time. Although there is some seasonality with peaks in particular months, it is obvious that buyers are more likely than ever to make purchases online.
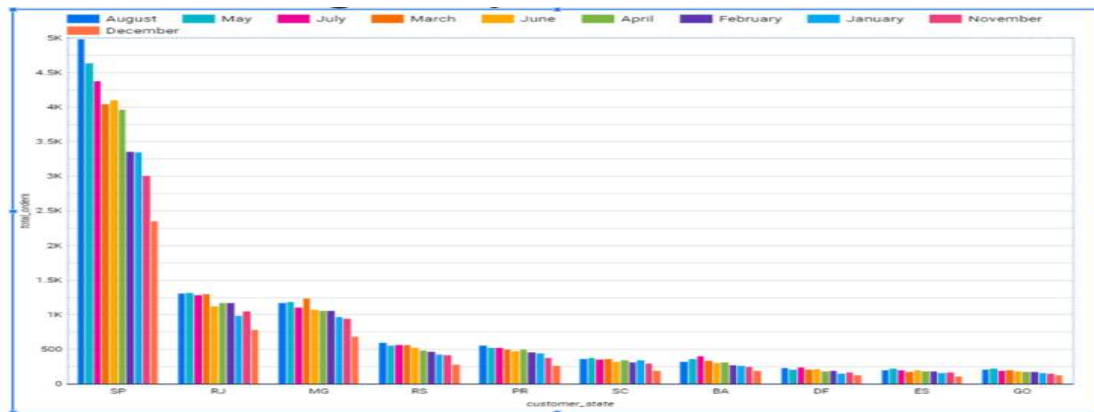
During August 2018 and September 2018, there was a significant decline, and it's possible that the cause was data noise. And the highest orders received in 2017-11 , and obviously the highest amount earned by the company is also in 2017-11.

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

1. **Get month on month orders by states::**

```sql
select month,customer_state, count(order_id) as count_orders from
(SELECT o.*,c.*,extract(year from o.order_purchase_timestamp) as year,
extract(month from o.order_purchase_timestamp) as month FROM
`sql-project-380605.Company.orders` as o  join
`sql-project-380605.Company.customers` as c on o.customer_id=c.customer_id ) as tbl
group by 2,1 order by 2,1
```


| Row	month	customer_state	count_orders
1	1	AC	8
2	2	AC	6
3	3	AC	4
4	4	AC	9
5	5	AC	10
6	6	AC	7
7	7	AC	9
8	8	AC	7
9	9	AC	5
10	10	AC	6

So,From results we can conclude that the highest number of orders occurred in 2018-08 for the state SP. And also Total of Customer Orders per state is highest for SP followed by RJ and MG states. Total of Customer Orders per state is lowest for RR followed by AP and AC states.

3)2)Distribution of customers across the states in Brazil::

```
select customer_state,count(distinct customer_id) as count_customers from
`sql-project-380605.Company.customers` group by 1 order  by 2 desc
```


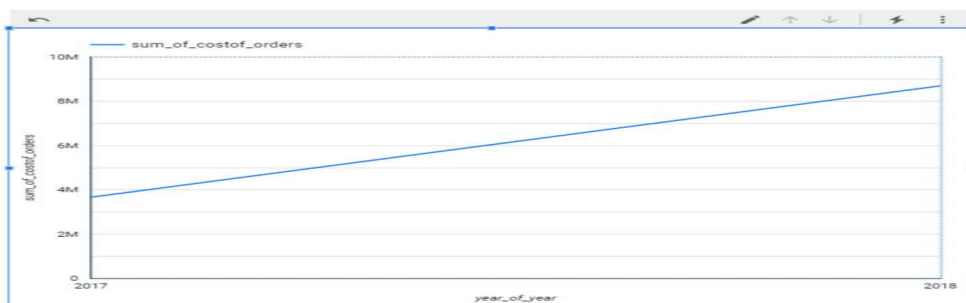| Row	customer_state	count_customers
1	SP	41746
2	RJ	12852
3	MG	11635
4	RS	5466
5	PR	5045
6	SC	3637
7	BA	3380
8	DF	2140
9	ES	2033
10	GO	2020
11	PE	1652
12	CE	1336
13	PA	975
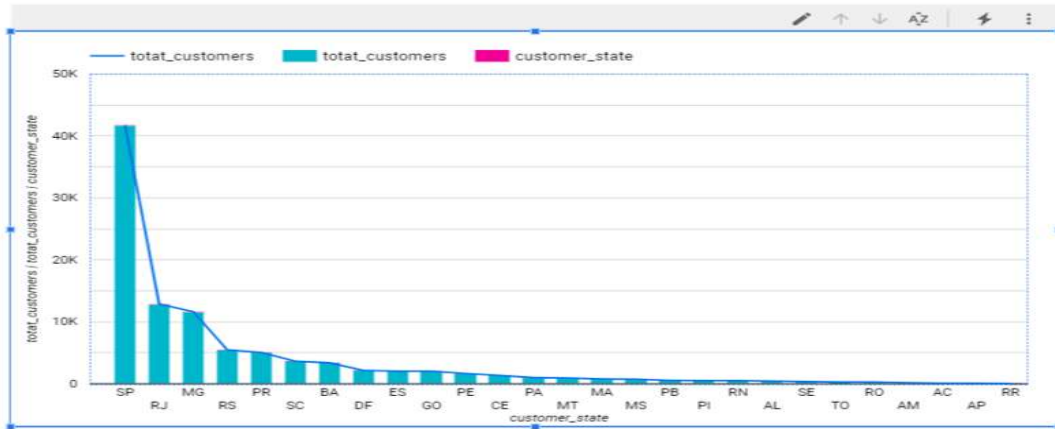
4)Impact on Economy: Analyse the money movement by e-commerce by looking at order prices, freight and others::

1. **Get % increase in cost of orders from 2017 to 2018 (include months between Jan to Aug only) - You can use "payment_value" column in payments table**

```
4)1)select distinct year, round(sum(sum_value)over(partition by year),2) as ind_sum
from(
select year,month ,round(sum(payment_value),2) as sum_value from
(SELECT o.*,p.*,extract(year from o.order_purchase_timestamp) as year,
extract(month from o.order_purchase_timestamp) as month FROM
`sql-project-380605.Company.orders` as o  join
`sql-project-380605.Company.payments` as p on o.order_id=p.order_id ) as tbl
group by 2,1 having month between 1 and 8 order by 2,1) as tb
```


So, we can conclude that the % increase in cost of orders from 2017 - 2018 between jan-aug is **+139.6%**.

4)2)Mean & Sum of price and freight value by customer state

```sql
select customer_state,sum_of_price,
round(tbl2.sum_of_price/tbl2.cnt_of_customers ,2) as mean_of_price_by_cust_state
,sum_of_freight_val,
round(tbl2.sum_of_freight_val/tbl2.cnt_of_customers,2) as
mean_of_freight_val_by_cust_state from(
select customer_state,count(customer_unique_id) as
cnt_of_customers,round(sum(price),2) as sum_of_price,ROUND(sum(freight_value),2) as
sum_of_freight_val from(
SELECT i.*,o.*,c.* FROM `sql-project-380605.Company.order_items` as i  join
`sql-project-380605.Company.orders` as o on i.order_id=o.order_id
join `sql-project-380605.Company.customers` as c on o.customer_id=c.customer_id )
as tbl
group by 1 order by 3 desc) as tbl2
```


Row	customer_state	sum_of_price	mean_of_price_by_cust_state	sum_of_freight_val	mean_of_freight_val_by_cust_state
1	SP	sum_of_price	109.65	718723.07	15.15
2	RJ	1824092.67	125.12	305589.31	20.96
3	MG	1585308.03	120.75	270853.46	20.63
4	RS	750304.02	120.34	135522.74	21.74
5	PR	683083.76	119.0	117851.68	20.53
6	SC	520553.34	124.65	89660.26	21.47
7	BA	511349.99	134.6	100156.68	26.36
8	DF	302603.94	125.77	50625.5	21.04
9	GO	294591.95	126.27	53114.98	22.77
10	ES	275037.31	121.91	49764.6	22.06

Results per page: 50   1 – 27 of 27

- It's fascinating to observe how some jurisdictions have a high overall volume of sales yet a low average order price. For instance, if we take a look at SP (São Paulo), we can see that it is the state with the most valuable e-commerce sales but it is also the state where clients pay less per order.
- Mean price by customer state is max for PB followed by AL,AC. And Mean price by customer state is min for SP followed by PR,RS.
- Sum price by customer state is max for SP followed by RJ,MG . And the sum price by customer state is min for RR.
- Mean of freight value by customer state is max for RR followed by PB and RO.And Mean of freight value by customer state is min for SP followed by PR and MG.
- Sum of freight value by customer state is max for SP followed by RJ, MG.And sum of freight value by customer state is min for RR.
- Customers in Roraima (RR), Paraba (PB), Rondônia (RO), and Acre (AC), for instance, typically pay more for freight than anyone else.

5)Analysis on sales, freight and delivery time::

5)5)Top 5 states with highest average freight value - sort in desc limit 5

```
select customer_state,mean_of_freight_val_by_cust_state from
(select
customer_state,sum_of_freight_val,round(tbl2.sum_of_freight_val/tbl2.cnt_of_custome
rs,2) as mean_of_freight_val_by_cust_state from
(select customer_state,count(customer_unique_id) as
cnt_of_customers,ROUND(sum(freight_value),2) as sum_of_freight_val
 from(SELECT i.*,o.*,c.* FROM `sql-project-380605.Company.order_items` as i  join
`sql-project-380605.Company.orders` as o on i.order_id=o.order_id
join `sql-project-380605.Company.customers` as c on o.customer_id=c.customer_id )
as tbl group by 1 order by 2 desc) as tbl2) as tbl3 order by 2 desc limit 5
```

Query results

	JOB INFORMATION	RESULTS	JSON	EXECUTION DETAILS
Row	customer_state			mean_of_freight_val_by_cust_state
1	RR			42.98
2	PB			42.72
3	RO			41.07
4	AC			40.07
5	PI			39.15

So,those top 5 states are RR followed by PB,RO,AC,PI.

5)5)Top 5 states with lowest average freight value - sort in ascending limit 5

```
select customer_state,mean_of_freight_val_by_cust_state
 from(select
customer_state,sum_of_freight_val,round(tbl2.sum_of_freight_val/tbl2.cnt_of_custome
rs,2) as mean_of_freight_val_by_cust_state
 from(select customer_state,count(customer_unique_id) as
cnt_of_customers,ROUND(sum(freight_value),2) as sum_of_freight_val
 from(SELECT i.*,o.*,c.* FROM `sql-project-380605.Company.order_items` as i  join
`sql-project-380605.Company.orders` as o on i.order_id=o.order_id
join `sql-project-380605.Company.customers` as c on o.customer_id=c.customer_id )
as tbl group by 1 order by 2 desc) as tbl2) as tbl3 order by 2  limit 5
```
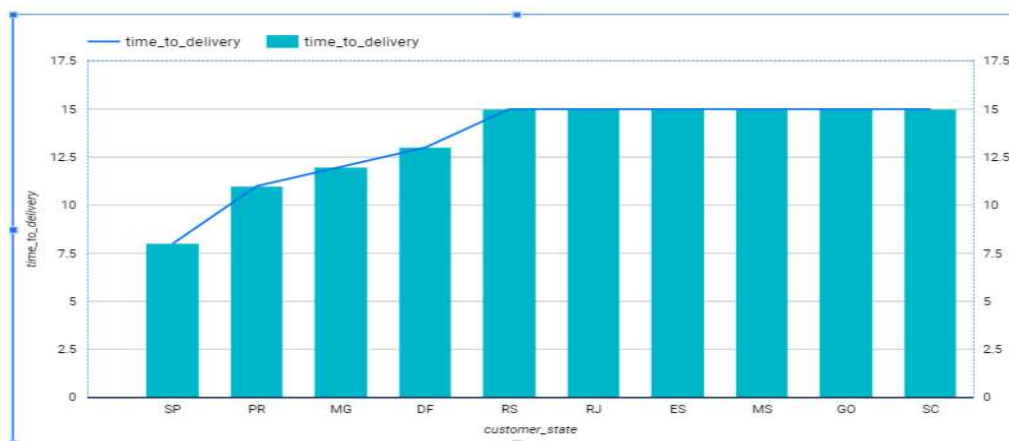
So,Top 5 states with lowest average freight value are SP followed by PR,MG,RJ,DF.

5)6)Top 5 states with lowest average time to delivery::

```sql
select customer_state,round(sum(time_to_delivery)/count(customer_unique_id),2) as
mean from(
SELECT o.*,c.*,date_diff(extract(date from
o.order_delivered_customer_date),extract(date from o.order_purchase_timestamp),day)
as time_to_delivery
 FROM `sql-project-380605.Company.orders` as o  join
`sql-project-380605.Company.customers` as c on o.customer_id=c.customer_id ) as tbl
group by 1 order by mean  limit 5
```


So,those top 5 states with lowest average time to delivery are SP followed by PR,MG,DF,SC.

5)6)Top 5 states with highest average time to delivery::

```
select customer_state,round(sum(time_to_delivery)/count(customer_unique_id),2) as
mean from(
SELECT o.*,c.*,date_diff(extract(date from
o.order_delivered_customer_date),extract(date from o.order_purchase_timestamp),day)
as time_to_delivery
 FROM `sql-project-380605.Company.orders` as o  join
`sql-project-380605.Company.customers` as c on o.customer_id=c.customer_id ) as tbl
group by 1 order by mean desc limit 5
```

Row	customer_state	cnt
1	AP	26.78
2	RR	26.15
3	AM	25.82
4	AL	23.55
5	PA	23.02

So,those top 5 states with highest average time to delivery are AP followed by RR,AM,AL,PA

5)7)Top 5 states where delivery is not so fast compared to estimated date::

```
Select
customer_state,round(sum(diff_estimated_delivery)/count(customer_unique_id),2) as
mean from( SELECT o.*,c.*,date_diff(extract(date from
o.order_estimated_delivery_date),extract(date from
o.order_delivered_customer_date),day) as diff_estimated_delivery
 FROM `sql-project-380605.Company.orders` as o  join
`sql-project-380605.Company.customers` as c on o.customer_id=c.customer_id ) as tbl
group by 1 order by mean limit 5
```

Row	customer_state	cnt
1	AL	8.37
2	MA	9.19
3	SE	9.59
4	ES	10.3
5	CE	10.34

So, Top 5 states where delivery is not so fast compared to estimated date are AL,MA,SE,ES,CE.

5)7)Top 5 states where delivery is really fast compared to estimated date::

```sql
select
customer_state,round(sum(diff_estimated_delivery)/count(customer_unique_id),2) as
mean from(
SELECT o.*,c.*,date_diff(extract(date from
o.order_estimated_delivery_date),extract(date from
o.order_delivered_customer_date),day) as diff_estimated_delivery
 FROM `sql-project-380605.Company.orders` as o  join
`sql-project-380605.Company.customers` as c on o.customer_id=c.customer_id ) as tbl
group by 1 order by mean desc limit 5
```

Query results

	JOB INFORMATION	RESULTS	JSON	E

Row	customer_state	cnt
1	AC	20.47
2	AP	19.4
3	RO	19.31
4	AM	19.17
5	RR	15.41

5)3)Group data by state, take mean of freight_value::

```sql
select
customer_state,sum_of_freight_val,round(tbl2.sum_of_freight_val/tbl2.cnt_of_custome
rs,2) as mean_of_freight_val_by_cust_state
 from(select customer_state,count(customer_unique_id) as
cnt_of_customers,ROUND(sum(freight_value),2) as sum_of_freight_val
 from(SELECT i.*,o.*,c.* FROM `sql-project-380605.Company.order_items` as i  join
`sql-project-380605.Company.orders` as o on i.order_id=o.order_id
join `sql-project-380605.Company.customers` as c on o.customer_id=c.customer_id )
as tbl group by 1 order by 2 desc) as tbl2
```

Query results

	JOB INFORMATION	RESULTS	JSON	EXECUTION DETAILS

Row	customer_state	sum_of_freight_	mean_of_freight
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	SC	89660.26	21.47
7	BA	100156.68	26.36
8	DF	50625.5	21.04
9	GO	53114.98	22.77
10	ES	49764.6	22.06
11	PE	59449.66	32.92

5)3)Group data by state, take mean of diff_estimated_delivery,
Mean of time_to_delivery::

```sql
select
customer_state,round(sum(diff_estimated_delivery)/count(customer_unique_id),2) as
mean_diff_estimated_delivery,round(sum(time_to_delivery)/count(customer_unique_id),
2) as mean_time_to_delivery from(
SELECT o.*,c.*,date_diff(extract(date from
o.order_estimated_delivery_date),extract(date from
o.order_delivered_customer_date),day) as
diff_estimated_delivery,date_diff(extract(date from
o.order_delivered_customer_date),extract(date from o.order_purchase_timestamp),day)
as time_to_delivery FROM `sql-project-380605.Company.orders` as o  join
`sql-project-380605.Company.customers` as c on o.customer_id=c.customer_id ) as tbl
group by 1
```

Query results

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

Row	customer_state	mean_diff_estimated_delivery	mean_time_to_delivery
1	RN	13.34	18.79
2	CE	10.34	20.3
3	RS	13.6	14.91
4	SC	11.22	14.54
5	SP	10.74	8.44
6	MG	12.92	11.66
7	BA	10.4	18.57
8	RJ	11.31	14.65
9	GO	11.81	15.05
10	MA	9.19	20.65

5)(1,2)Find time_to_delivery & diff_estimated_delivery for each customer::

```sql
select customer_unique_id,time_to_delivery,diff_estimated_delivery from (
SELECT o.*,c.*,date_diff(extract(date from
o.order_estimated_delivery_date),extract(date from
o.order_delivered_customer_date),day) as
diff_estimated_delivery,date_diff(extract(date from
o.order_delivered_customer_date),extract(date from o.order_purchase_timestamp),day)
as time_to_delivery FROM `sql-project-380605.Company.orders` as o  join
`sql-project-380605.Company.customers` as c on o.customer_id=c.customer_id) as tbl
```

Query results

JOB INFORMATION	RESULTS	JSON	EXECUTION DETAILS	EXECUTION GI

| Row	customer_unique_id	time_to_delivery	diff_estimated_delivery
1	402cce5c0509000eed9e77fec...	46	-14
2	6ba00666ab7eada5ceec279b2...	48	-16
3	8f399f3b7ace8e6245422c9e1f...	37	-6
4	5bed15735f2d1bdfb2ba82869...	47	-28
5	3306ce16186d1c3938adbfc02...	39	21
6	df5e44ce11714dec60482a8bd...	34	0
7	30d13c3e324acc4ee50bda3e4...	31	-9
8	7804bb80e888801c61290f838...	44	-17
9	b80affc9a7a9d034cd9045ecc3...	30	-2
10	3f1ae9854823f7c0a8027bd99...	37	-15
11	236cb33406610c4551abb79a9...	39	-13

6) Payment Type Analysis::

6)1)Month over Month count of orders for different payment types:

```sql
select year,month,payment_type,count(customer_id) as count_of_orders from(
SELECT o.*,p.*,extract(year from o.order_purchase_timestamp) as year, extract(month
from o.order_purchase_timestamp) as month FROM `sql-project-380605.Company.orders`
as o  join `sql-project-380605.Company.payments` as p on o.order_id=p.order_id ) as
tbl group by 1,2,3 order by 4 desc
```

Query results

JOB INFORMATION	RESULTS	JSON	EXECUTION DETAILS	EXECU

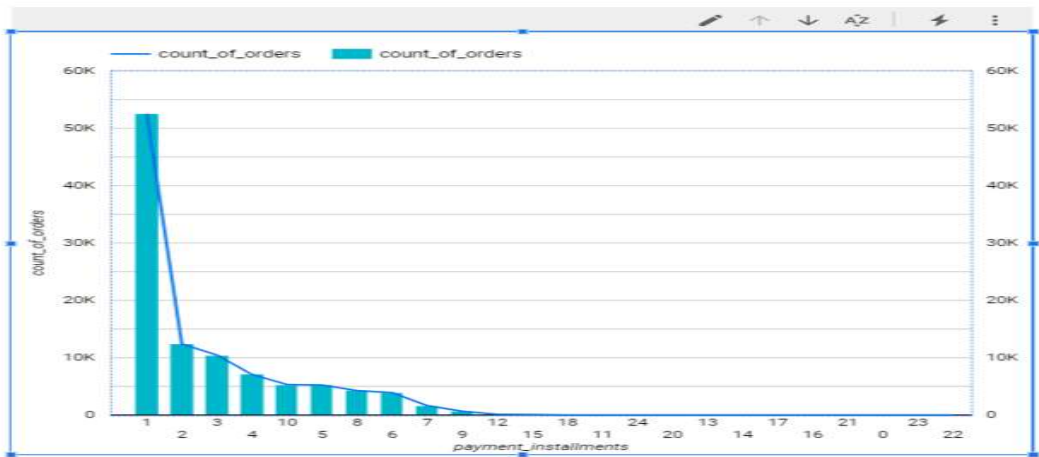| Row	year	month	payment_type	count_of_orders
1	2017	11	credit_card	5897
2	2018	3	credit_card	5691
3	2018	1	credit_card	5520
4	2018	5	credit_card	5497
5	2018	4	credit_card	5455
6	2018	2	credit_card	5253
7	2018	8	credit_card	4985
8	2018	6	credit_card	4813
9	2018	7	credit_card	4755
10	2017	12	credit_card	4377
11	2017	10	credit_card	3524
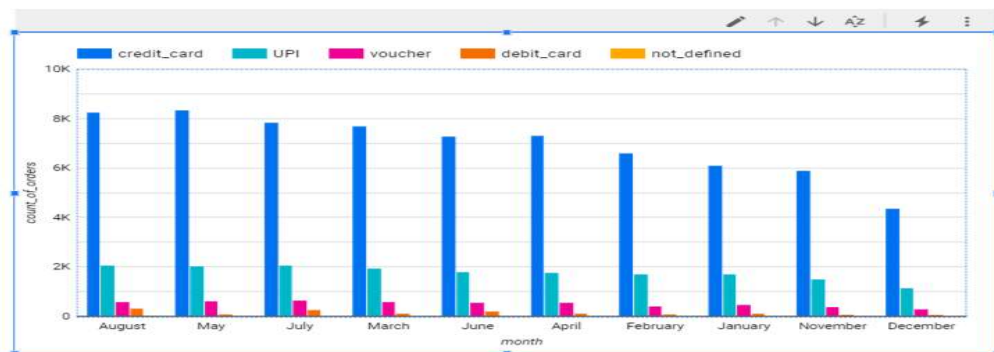
We can see that credit card payments accounted for a significant portion of Brazilian online sales. In addition, beginning 2018-03, a slight decline in this form of payment may be observed. On the other hand, a growing trend in debit card payments has been seen since 2018-05, which presents a good potential for investors to enhance services for these types of payments.

6)2)Count of orders based on the no. of payment instalments

```
select payment_installments ,count(order_id) as count_of_orders from
`sql-project-380605.Company.payments` group by 1 order by 2 desc
```

Query results

	JOB INFORMATION	RESULTS	JSON

Row	payment_installments	count_of_orders
1	1	52546
2	2	12413
3	3	10461
4	4	7098
5	10	5328
6	5	5239
7	8	4268
8	6	3920
9	7	1626
10	9	644

We can see that Brazilian clients prefer to pay for their orders in one instalment rather than several, and it's important to note the number of payments made in ten instalments.

Actionable Insights:
- We can get details of the states and cities from where the orders were placed.
- We can also get sales in each state every month .
- We can also get the information of total sales and orders in each month common in all years across all the states
- From this we can say in which people are more tending to buy
- We can also get the information on which time of the day people are tending to order , as a result we got to know that people tend to buy more in the afternoon followed by night , morning, and dawn.
- We can also know total no.of customers in each state from this we can get which state we have more customers i.e in state SP we have more customers.
- We can also know the increase in percent of sales in a particular time period for ex: in the time period given in 4 th the increase in sales is 139% .
- We can also find information like which payment mode people are preferring to use i.e Credit card .
- We can also get the information on no.of instalments people are tend to buy the product i.e. instalment one

Recommendations:
- We learned that in the month of December the orders are reduced so it is recommended to have more offers and some special day sales in the month of December .
- We got to know that sales in the afternoon period are more so it was recommended that if the platform wanted to give any advertisements it was better to give in this time period.
- We also get to known that the customers in the states like PB,RN,SE,RO are very less so it was recommended to display the products and offers that attract the people in that specific community
- We also get to know that in the states like RR, AP ,AM,AL,PA the no.of days taken to deliver the product is more so it was recommended to improve the transportation to deliver the product faster.
- We get to know that in the states like PB,RR,RO,AC,PI.
- The average freight height is more so it was recommended to reduce the charges of delivery to attract the customers by utilising the transportation to the best extent.
- We can get to know that in the states like AL,MA,SE,RR,AP. The delivery date was late when compared to estimated delivery which may lead to loss of interest on the platform to the customers so it recommended that to decrease the no.of days to deliver the product when compared to estimated delivery date if it is not possible then it recommended that to increase the estimated delivery at least.
- We also get to know that people prefer to pay in one instalment.