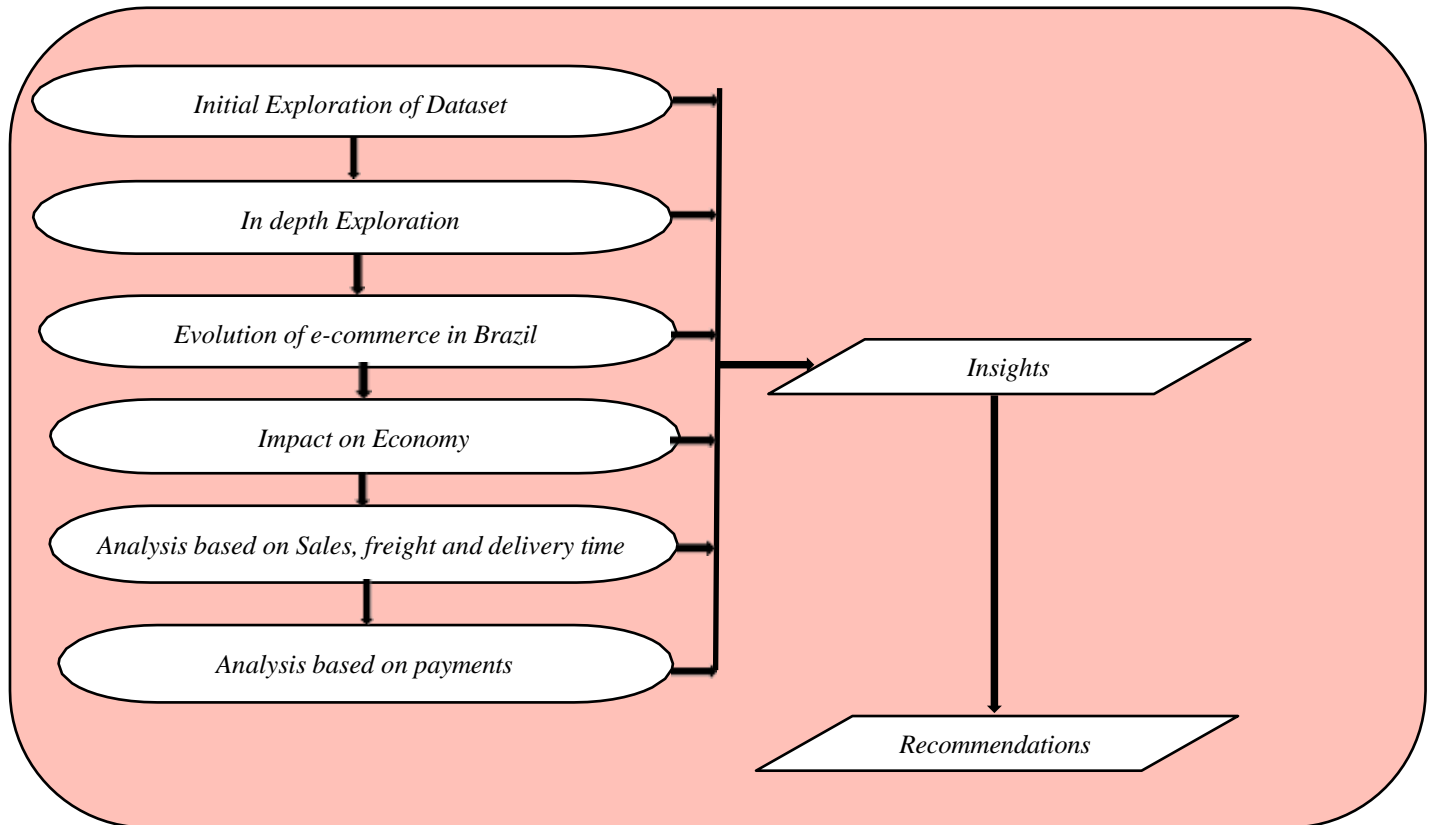*Priyanshu Raj*

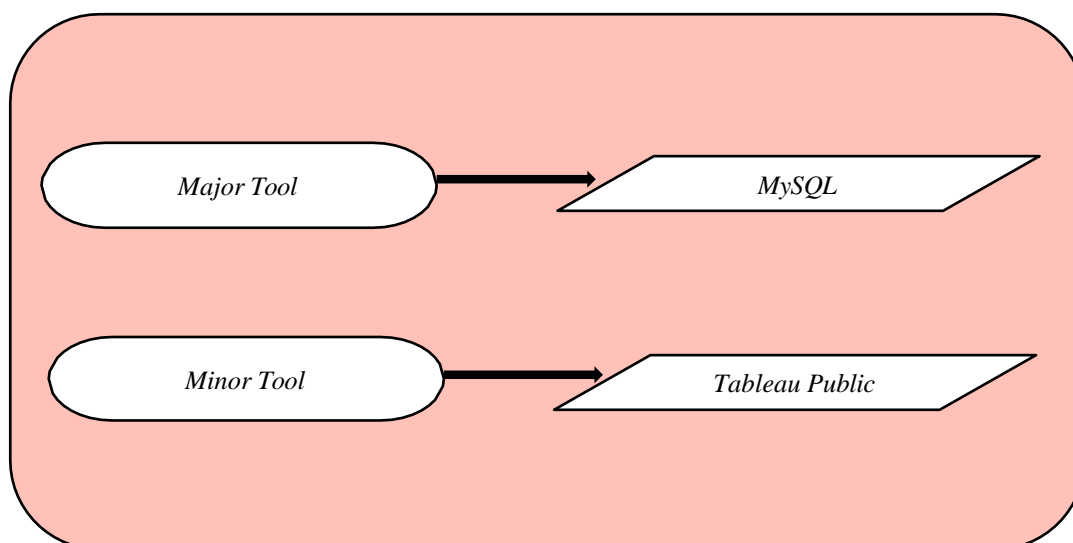*priyanshuraj5959@gmail.com*

Unveiling E-commerce Brand Dynamics

Description of Business case:

The dataset belongs to renowned retailer brand in United States. The company makes itself a preferred shopping destination by offering outstanding value, inspiration, innovation and exceptional user experience that no other retailer can offer. This Business case focuses company's operation in Brazil. By analyzing this dataset, it might be possible to gain valuable insights about company's operation and can shed light on various aspects of business.

Data Exploration: The analysis is performed with given below structure in flowchart

Tools used:

1. Initial Exploration of Dataset

> describe company.customers   #similar query is used to for every tables and made a logical schema.

1.2. What is the start date and end of available dataset?
Query: 1.2

> *select*
>   *min(order_purchase_timestamp) as start_date,*
>   *max(order_purchase_timestamp) as end_date*
> *from company.orders;*

Result:

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

1.3. What is the duration of time of available dataset (in months)?
Query: 1.3

> select
>   datediff(max(order_purchase_timestamp),min(order_purchase_timestamp))/30 as no_of_month
> from
>   company.orders

Result:

no_of_month
▶ 25.7667

1.4. find out the total number of customers in Brazil?
Query:1.4

```sql
select
        count(distinct customer_id) as total_customers
from
        company.customers
```

Result:

total_customers
▶ 99441

1.5. find out the top states with highest percentage of customers in Brazil?
Query:1.5

```sql
select
        customer_state,
        round(count(customer_state)*100/(select count(*) from company.customers)) as percentage_of_customer_base
from
        company.customers
group by customer_state
order by 2 desc;
```

Result:

customer_state	percentage_of_customer_base
▶ SP	42
RJ	13
MG	12
PR	5
RS	5
SC	4
BA	3
GO	2
ES	2
DF	2
PE	2
PA	1
MA	1
MS	1
CE	1
MT	1
PB	1
RN	0
AM	0
AP	0

2. In-Depth Exploration

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

```
select
        extract(year from order_purchase_timestamp) as year,
        extract(month from order_purchase_timestamp) as month,
        count(order_id) as order_receieved
from
        company.orders
group by year,month
order by year,month;
```

Result:

year	month	order_receieved
2016	9	4
2016	10	324
2016	12	1
2017	1	800
2017	2	1780
2017	3	2682
2017	4	2404
2017	5	3700
2017	6	3245
2017	7	4026
2017	8	4331
2017	9	4285
2017	10	4631
2017	11	7544
2017	12	5673
2018	1	7269
2018	2	6728
2018	3	7211
2018	4	6939
2018	5	6873

Result 15 ×

Note:

1. it may be noted that there is increase in the order_recevied values as time have passed. It implies that there is a clear trend of growing business in Brazil. We used tableau for even better visualization as given below.
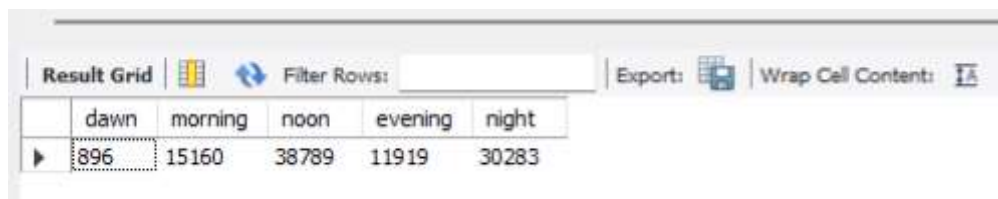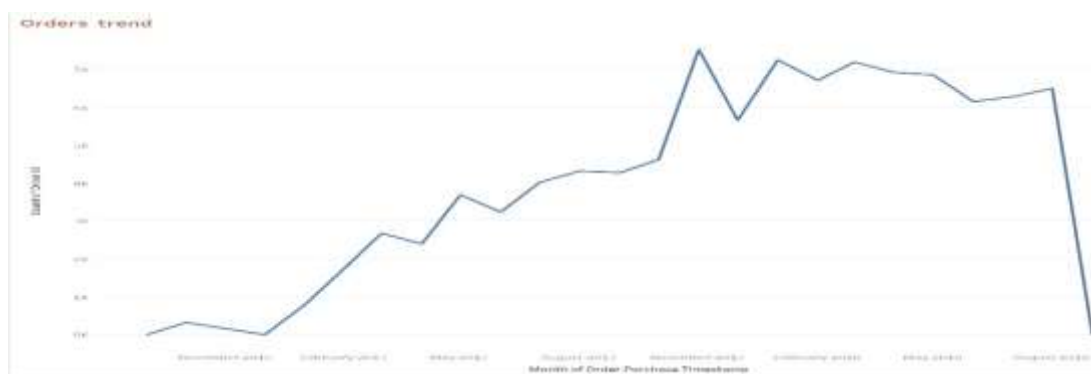

* Trend is captured using Tableau public

2.3 let us try to check at what part of day, do Brazilians place order. To analyze the same let us frame a rubric as given below

Category	Time range
Dawn	5:00-6:59
Morning	7:00-10:59
Noon	11:00-16:59
Evening	17:00-18:59
Night	19:00-4:59

Query: 2.3

```
select
        sum(case when extract(hour from order_purchase_timestamp) between 4 and 6 then 1 else 0 end) as "dawn",
        sum(case when extract(hour from order_purchase_timestamp) between 7 and 10 then 1 else 0 end) as "morning",
        sum(case when extract(hour from order_purchase_timestamp) between 11 and 16 then 1 else 0 end) as "noon",
        sum(case when extract(hour from order_purchase_timestamp) between 17 and 18 then 1 else 0 end) as "evening",
        sum(case when extract(hour from order_purchase_timestamp) between 19 and 0 then 1 else 0 end) +
        sum(case when extract(hour from order_purchase_timestamp) between 1 and 3 then 1 else 0 end) as "night"
from company.orders
```

Result:

dawn	morning	noon	evening	night
896	15160	38789	11919	30283

Note:

1). It may be observed that Brazilians have tendency to place more orders at noon followed by night.
2). Minimum orders are placed dawn and morning.

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

3.1 What are the number of orders place month on month basis for each state?
Query: 3.1

```sql
select
        customer_state,
        extract(year from order_purchase_timestamp) as year,
        extract(month from order_purchase_timestamp) as month,
        count(order_id) as no_of_orders
from
        company.orders o
        join company.customers c on c.customer_id=o.customer_id
group by
        customer_state,year,month
order by
        1,2,3
```

Result:

customer_state	year	month	no_of_orders
AC	2017	1	2
AC	2017	2	3
AC	2017	3	2
AC	2017	4	5
AC	2017	5	8
AC	2017	6	4
AC	2017	7	5
AC	2017	8	4
AC	2017	9	5
AC	2017	10	6
AC	2017	11	5

Result 22 ✕

We may visualize same output using tableau as given below:
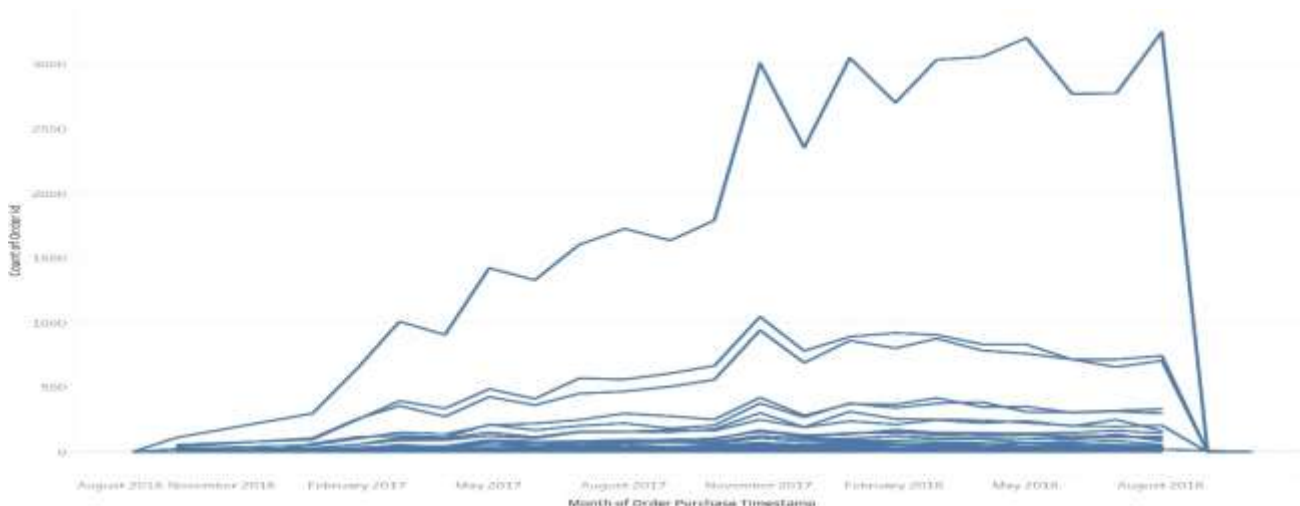
Insights:
1. Every line in the curve represents a state of Brazil.
2. It may be noted that up to November 2017, there is a rising trend whereas after that it is getting saturated. Although data is for small duration of time so further investigation is required for confirmation.
3. It may also be observed that November month receives a spike across all the states.

3.2 how many states in Brazil have engaged business with the company?
Query: 3.2

```
select
        count(distinct customer_state) as no_of_states
from company.customers
```

Result:

Result Grid	Filter Rows:	Export:	Wrap Cell Content:

no_of_states
▶ 27

3.3 how many cities in Brazil have engaged business with the company?
Query: 3.3

```
select
        count(distinct customer_city) as no_of_cities
from company.customers
```

Result:

Result Grid	Filter Rows:	Export:	Wrap Cell Content:

no_of_cities
▶ 4119

3.4 Write a query to find out percentage of customer base across various states
Query: 3.4

```
select
        customer_city,
        customer_state,
        (100*count(customer_id)/(select count(*) from company.customers)) as percentage_of_customerbase
from company.customers
group by customer_state,customer_city
order by percentage_of_customerbase desc
```

Result:

customer_city	customer_state	percentage_of_customerbase
▶ sao paulo	SP	15.6274
rio de janeiro	RJ	6.9207
belo horizonte	MG	2.7886
brasilia	DF	2.1430
curitiba	PR	1.5296
campinas	SP	1.4521
porto alegre	RS	1.3868
salvador	BA	1.2520
guarulhos	SP	1.1957
sao bernardo do campo	SP	0.9433
niteroi	RJ	0.8538

Result 4 ✕

Insights:
1. Company a presence across all the 27 states in Brazil.
2. Company is engaged in business across 4119 cities
3. Company's biggest customer base in Brazil is in Sau Paulo followed by Rio de Jeneiro and Belo horizonte

4. Impact on Economy

4.1. what is the percentage rise of cost of orders from 2017 to 2018, considering business only between January to September as we don't have sufficient data for whole year.
Query: 4.1

```sql
with cte as
    (
        SELECT
            distinct YEAR(o.order_purchase_timestamp) AS year,
            SUM(p.payment_value) OVER (PARTITION BY YEAR(o.order_purchase_timestamp)) AS total_payment
        FROM company.orders o
        JOIN company.payments p ON p.order_id = o.order_id
        where month(o.order_purchase_timestamp) between 1 and 8
    )

select
    lag(total_payment,1,0) over() as 2017_total_payment,
    total_payment AS "2018_total_payment",
    ifnull(100*((total_payment-lag(total_payment,1,0) over())/lag(total_payment,1,0) over()),0) as percentage_rise_of_business
from cte
```

Result:

2017_total_payment	2018_total_payment	percentage_rise_of_business
3669022.1199999577	8694733.839999836	136.97687164665925

4.2. Calculate total and average value of order price for each state
Query: 4.2

```sql
Select
    c.customer_state,
    sum(p.payment_value) as total_order_price,
    avg(p.payment_value) as avg_order_price
from
    company.customers c
    join company.orders o on c.customer_id=o.customer_id
    join company.payments p on p.order_id=o.order_id
group by
    c.customer_state
order by
    2 desc,3 desc
```

Result:

customer_state	total_order_price	avg_order_price
SP	5998226.959999805	137.5046297739628
RJ	2144379.6899999995	158.52588822355287
MG	1872257.2600000077	154.70643364733166
RS	890898.5399999958	157.18040578687294
PR	811156.3799999995	154.15362599771942
SC	623086.4299999996	165.9793367075119
BA	616645.8200000026	170.81601662049934
DF	355141.07999999984	161.13479128856616
GO	350092.3100000004	165.7634043560608
ES	325967.5500000016	154.7069530137644
PE	324850.43999999925	187.99215277777733

4.3. Calculate total and average value of order freight for each state.
Query: 4.3

```sql
select
        c.customer_state,
        sum(ot.freight_value) as total_freight,
        avg(ot.freight_value) as avg_freight
from

        company.customers  c
        join company.orders o on c.customer_id=o.customer_id
        join company.order_items ot on o.order_id=ot.order_id
group by
        c.customer_state
order by
        2 desc,3 desc
```

Result:

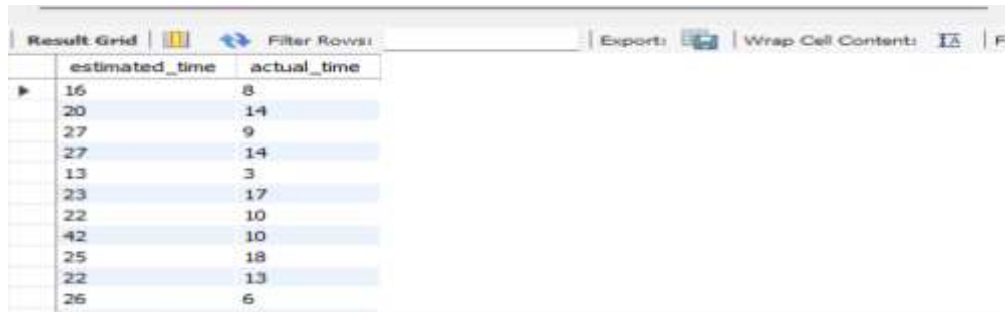| customer_state	total_freight	avg_freight
SP	718723.0699999852	15.147275390418875
RJ	305589.30999999796	20.960923931682416
MG	270853.4600000028	20.630166806306864
RS	135522.74000000235	21.735804330393318
PR	117851.68000000104	20.531651567944433
BA	100156.67999999903	26.363958936562
SC	89660.26000000015	21.470368773946397
PE	59449.659999999894	32.91786267995565
GO	53114.97999999994	22.7668152259322733
DF	50625.49999999984	21.041354945968347
ES	49764.5999999998	22.058776595744593

5. Analysis on sales, freight and delivery time

5.1 Find out time taken to deliver the order and estimated time
Query: 5.1

```
select
        datediff(order_estimated_delivery_date,order_purchase_timestamp) as estimated_time,
        datediff(order_delivered_customer_date, order_purchase_timestamp) as actual_time
from
        company.orders
where
        datediff(order_estimated_delivery_date,order_purchase_timestamp) is not null
        and
        datediff(order_delivered_customer_date, order_purchase_timestamp) is not null
```
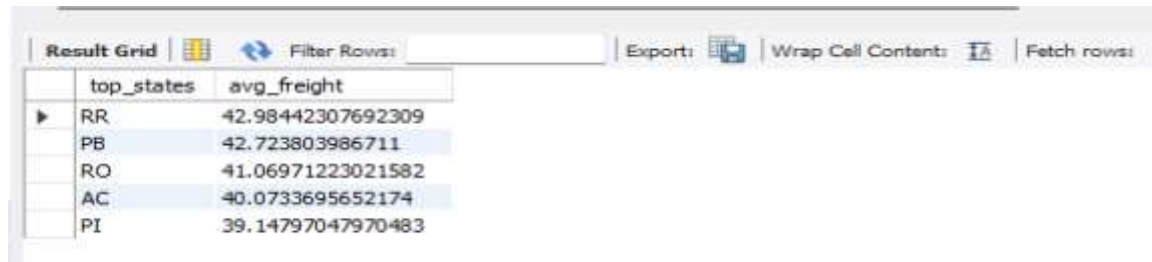
Result:

estimated_time	actual_time
16	8
20	14
27	9
27	14
13	3
23	17
22	10
42	10
25	18
22	13
26	6

5.2 What are the top 5 states with highest average freight value?
Query: 5.2

```
select
        c.customer_state as top_states,
        avg(ot.freight_value) as avg_freight
from
        company.customers c
        join company.orders o on c.customer_id=o.customer_id
        join company.order_items ot on o.order_id=ot.order_id
group by
        c.customer_state
order by 2 desc
limit 5
```

Result:

top_states	avg_freight
RR	42.98442307692309
PB	42.723803986711
RO	41.069712230021582
AC	40.0733695652174
PI	39.14797047970483

5.3 What are the top 5 states with lowest average freight value?
Query: 5.3

```
select
        c.customer_state as bottom_states,
        avg(ot.freight_value) as avg_freight
from
        company.customers c
        join company.orders o on c.customer_id=o.customer_id
        join company.order_items ot on o.order_id=ot.order_id
group by
        c.customer_state
order by 2 asc
limit 5
```

Result:

bottom_states	avg_freight
SP	15.147275390418875
PR	20.531651567944433
MG	20.630166806306864
RJ	20.960923931682416
DF	21.041354945968347

5.4 What are the top 5 states with highest delivery time?
Query: 5.4

```sql
select
        c.customer_state as top_states,
        avg(datediff(order_delivered_customer_date, order_purchase_timestamp)) as average_delivery_time
from
        company.orders o
        join company.customers c on c.customer_id=o.customer_id

group by c.customer_state
order by 2 desc
limit 5
```

Result:

top_states	average_delivery_time
RR	29.3415
AP	27.1791
AM	26.3586
AL	24.5013
PA	23.7252

5.5 What are the top 5 states with lowest delivery time?
Query: 5.5

```sql
select
        c.customer_state as bottom_states,
        avg(datediff(order_delivered_customer_date, order_purchase_timestamp)) as average_delivery_time
from
        company.orders o
        join company.customers c on c.customer_id=o.customer_id
group by c.customer_state
order by 2 asc
limit 5
```

Result:

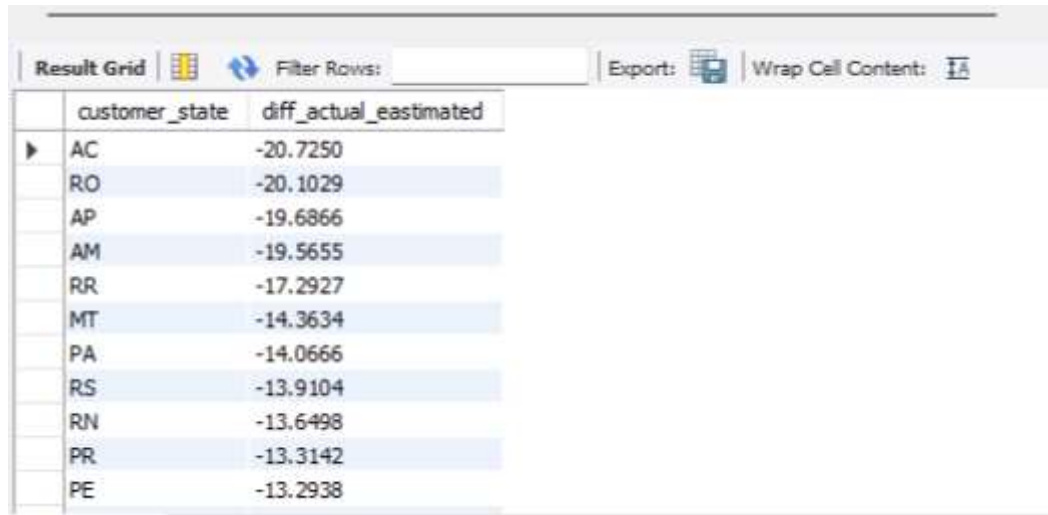| bottom_states	average_delivery_time
SP	8.7005
PR	11.9380
MG	11.9465
DF	12.8990
SC	14.9075

5.6 What are the top 5 states with fastest delivery?
Note: here fast delivery will be decided on difference between estimated and actual time take to delivery.

Query: 5.6

```
select
        c.customer_state,
        avg(datediff(order_delivered_customer_date, order_purchase_timestamp)-
        datediff(order_estimated_delivery_date,order_purchase_timestamp)) as diff_actual_eastimated

from
        company.orders o
        join company.customers c on c.customer_id=o.customer_id
group by c.customer_state
order by 2 asc
```

Result:

customer_state	diff_actual_eastimated
AC	-20.7250
RO	-20.1029
AP	-19.6866
AM	-19.5655
RR	-17.2927
MT	-14.3634
PA	-14.0666
RS	-13.9104
RN	-13.6498
PR	-13.3142
PE	-13.2938

5.7 What are the top 5 states with slowest delivery?
Note: here fast delivery will be decided on difference between estimated and actual time take to delivery.
Query: 5.7

```
select
        c.customer_state,
        avg(datediff(order_delivered_customer_date, order_purchase_timestamp)-
    datediff(order_estimated_delivery_date,order_purchase_timestamp)) as diff_actual_eastimated

from
company.orders o
join company.customers c on c.customer_id=o.customer_id
group by c.customer_state
order by 2 desc
```

Result:

customer_state	diff_actual_eastimated
AL	-8.7078
MA	-9.5718
SE	-10.0209
ES	-10.4962
BA	-10.7945
CE	-10.8045
MS	-11.0528
SP	-11.0764
PI	-11.3067
SC	-11.5083
RJ	-11.7669

6. Payment type Analysis

6.1. Find out no of orders placed month on month using different payment type.
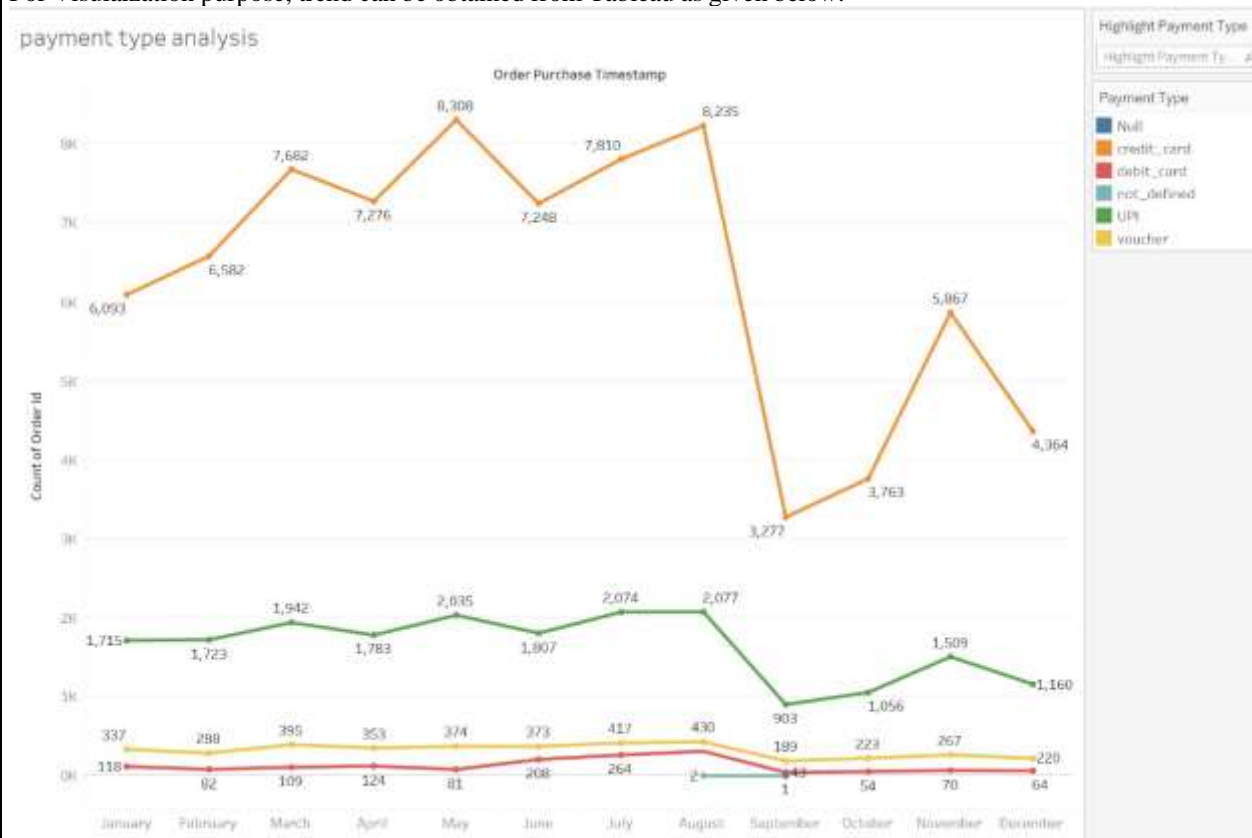
Query: 6.1

```
select
        payment_type,
        year(order_purchase_timestamp) as year,
        month(order_purchase_timestamp) as month,
        count(o.order_id) as no_of_orders
from
        company.orders o
        join company.payments p on o.order_id=p.order_id
group by
        payment_type,year,month
order by
        payment_type,year asc,month asc
```

Result:

payment_type	year	month	no_of_orders
credit_card	2016	9	3
credit_card	2016	10	254
credit_card	2016	12	1
credit_card	2017	1	583
credit_card	2017	2	1356
credit_card	2017	3	2016
credit_card	2017	4	1846
credit_card	2017	5	2853
credit_card	2017	6	2463
credit_card	2017	7	3086
credit_card	2017	8	3284

For Visulaization purpose, trend can be obtained from Tableau as given below:

6.2: find the number of orders placed on the basis of payment installments that have been paid.

Query: 6.2

```sql
select
        payment_installments,
        count(order_id) as no_of_orders
from
        company.payments
group by
        payment_installments
order by
        payment_installments  asc
```

Result:

payment_installments	no_of_orders
0	2
1	52546
2	12413
3	10461
4	7098
5	5239
6	3920
7	1626
8	4268
9	644
10	5328

7. Actionable Insights

1. We have only 26 months long dataset.
2. State with biggest customer base in Brazil is SP followed by RJ and MG with 42%,13% and 12% respectively.
3. There is a clear rising trend of online business in Brazil.
4. Brazilian people have tendency to place orders during noon followed by night when compared to any other time of the day.
5. The company is doing business across all the 27 states in Brazil.
4. The company have a wide presence across 4119 cities in Brazil.
5. Company's biggest customer base in Brazil is in Sau Paulo followed by Rio de Jeneiro and Belo horizonte
6. Number of orders increased by 137% from 2017 to 2018
7. SP, RJ and MG are the states giving maximum business. It must be noted that these same states also have least value of freight per order, which makes business more profitable in these states.
8. The states with highest average freight is RR followed by PB and RO
9. The states with lowest average freight is SP followed by PR and MG
10. The states with highest delivery time is RR followed by AP and AM
11. The states with lowest delivery time is SP followed by PR and MG
12. Brazilian people prefer credit card followed by UPI when compared to any other mode of payment.
13. Brazilian people prefer payment in 1 installment itself followed by 2 and 3 installments.

8. Recommendations

1. Most of the buyers use Credit Cards as buying medium so Company may look for tie ups with Banks to accommodate offers in Credit Cards.
2. UPI is second most preferred mode of payment and is emerging technology. Company may explore to provide offers to customer to attract new customers.
3. SP, RJ, MG are the states of top priority, they give most business with least avg freight. Company should manage resources in these states especially during festive season.
4. Market is in growing phase so company may look for expansion.