Business case Study for MNC

By Shankara N G

CompanyX is a globally renowned brand and a prominent retailer in the United States. CompanyX makes itself a preferred shopping destination by offering outstanding value, inspiration, innovation and an exceptional guest experience that no other retailer can deliver.

This particular business case focuses on the operations of CompanyX in Brazil and provides insightful information about 100,000 orders placed between 2016 and 2018. The dataset offers a comprehensive view of various dimensions including the order status, price, payment and freight performance, customer location, product attributes, and customer reviews.

By analyzing this extensive dataset, it becomes possible to gain valuable insights into CompanyX's operations in Brazil. The information can shed light on various aspects of the business, such as order processing, pricing strategies, payment and shipping efficiency, customer demographics, product characteristics, and customer satisfaction levels.

____

Dataset: https://drive.google.com/drive/folders/1TGEc66YKbD443nslRi1bWgVd238gJCnb

The data is available in 8 csv files:

1. **customers.csv**
2. **sellers.csv**
3. **order_items.csv**
4. **geolocation.csv**
5. **payments.csv**
6. **reviews.csv**
7. **orders.csv**
8. **products.csv**

____

The column description for these csv files is given below.

The customers.csv contain following features:

Features	Description
customer_id	ID of the consumer who made the purchase
customer_unique_id	Unique ID of the consumer
customer_zip_code_prefix	Zip Code of consumer's location
customer_city	Name of the City from where order is made
customer_state	State Code from where order is made (Eg. são paulo - SP)

The sellers.csv contains following features:

Features	Description
seller_id	Unique ID of the seller registered
seller_zip_code_prefix	Zip Code of the seller's location
seller_city	Name of the City of the seller
seller_state	State Code (Eg. são paulo - SP)

The order_items.csv contain following features:

Features	Description
order_id	A Unique ID of order made by the consumers
order_item_id	A Unique ID given to each item ordered in the order
product_id	A Unique ID given to each product available on the site
seller_id	Unique ID of the seller registered in CompanyX
shipping_limit_date	The date before which the ordered product must be shipped
price	Actual price of the products ordered
freight_value	Price rate at which a product is delivered from one point to another

The geolocations.csv contain following features:

Features	Description
geolocation_zip_code_prefix	First 5 digits of Zip Code
geolocation_lat	Latitude
geolocation_lng	Longitude

geolocation_city	City
geolocation_state	State

The payments.csv contain following features:

Features	Description
order_id	A Unique ID of order made by the consumers
payment_sequential	Sequences of the payments made in case of EMI
payment_type	Mode of payment used (Eg. Credit Card)
payment_installments	Number of installments in case of EMI purchase
payment_value	Total amount paid for the purchase order

The orders.csv contain following features:

Features	Description
order_id	A Unique ID of order made by the consumers
customer_id	ID of the consumer who made the purchase
order_status	Status of the order made i.e. delivered, shipped, etc.
order_purchase_timestamp	Timestamp of the purchase
order_delivered_carrier_date	Delivery date at which carrier made the delivery
order_delivered_customer_date	Date at which customer got the product
order_estimated_delivery_date	Estimated delivery date of the products

The reviews.csv contain following features:

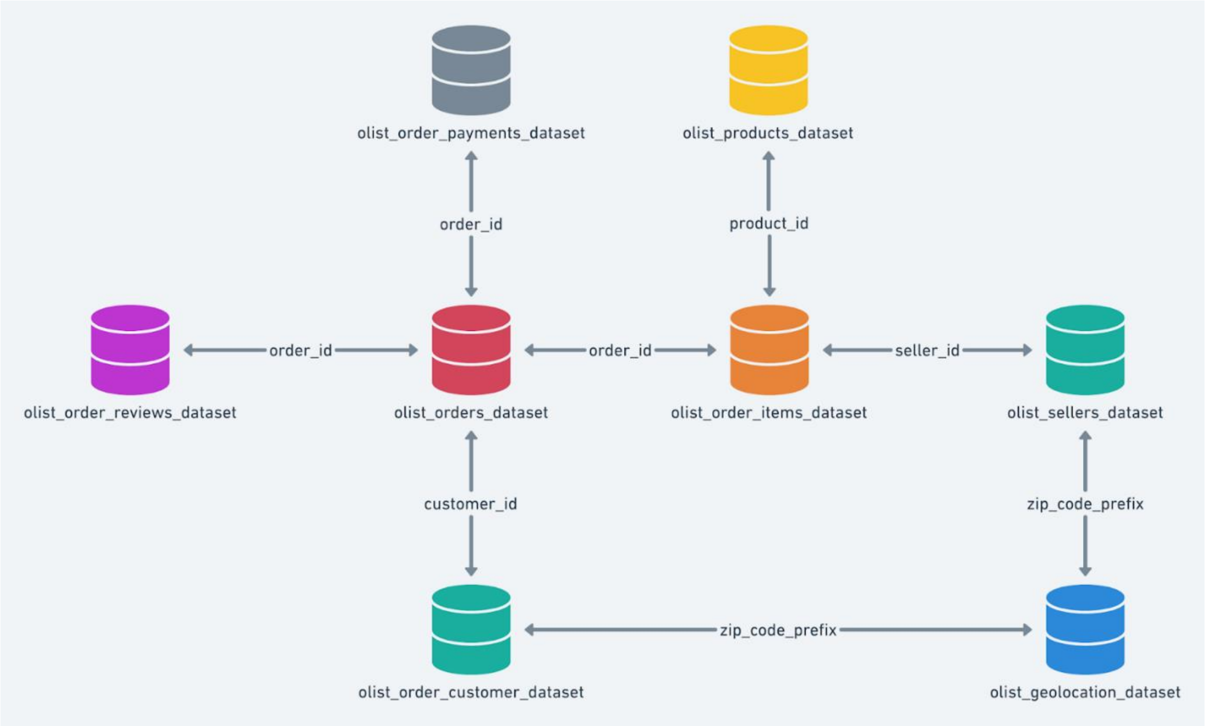
Features	Description
review_id	ID of the review given on the product ordered by the order id
order_id	A Unique ID of order made by the consumers
review_score	Review score given by the customer for each order on a scale of 1-5
review_comment_title	Title of the review
review_comment_message	Review comments posted by the consumer for each order
review_creation_date	Timestamp of the review when it is created

review_answer_timestamp	Timestamp of the review answered

The products.csv contain following features:

Features	Description
product_id	A Unique identifier for the proposed project.
product_category_name	Name of the product category
product_name_lenght	Length of the string which specifies the name given to the products
product_description_lenght	Length of the description written for each product ordered on the s
product_photos_qty	Number of photos of each product ordered available on the shoppi
product_weight_g	Weight of the products ordered in grams
product_length_cm	Length of the products ordered in centimeters
product_height_cm	Height of the products ordered in centimeters
product_width_cm	Width of the product ordered in centimeters

Dataset schema:



Problem Statement:

Assuming you are a data analyst/ scientist at CompanyX, you have been assigned the task of analyzing the given dataset to extract valuable insights and provide actionable recommendations.

What does 'good' look like?

1. **Import the dataset and do usual exploratory analysis steps like checking the structure & characteristics of the dataset:**
    1. **Data type of all columns in the "customers" table.**
    2. **Get the time range between which the orders were placed.**
    3. **Count the Cities & States of customers who ordered during the given period.**
2. **In-depth Exploration:**
    1. **Is there a growing trend in the no. of orders placed over the past years?**
    2. **Can we see some kind of monthly seasonality in terms of the no. of orders being placed?**
    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**
3. **Evolution of E-commerce orders in the Brazil region:**
    1. **Get the month on month no. of orders placed in each state.**
    2. **How are the customers distributed across all the states?**
4. **Impact on Economy: Analyze the money movement by e-commerce by looking at order prices, freight and others.**
    1. **Get the % increase in the cost of orders from year 2017 to 2018 (include months between Jan to Aug only).**
       You can use the "payment_value" column in the payments table to get the cost of orders.
    2. **Calculate the Total & Average value of order price for each state.**
    3. **Calculate the Total & Average value of order freight for each state.**
5. **Analysis based on sales, freight and delivery time.**

1. **Find the no. of days taken to deliver each order from the order's purchase date as delivery time.**
   Also, calculate the difference (in days) between the estimated & actual delivery date of an order.
   Do this in a single query.

   You can calculate the delivery time and the difference between the estimated & actual delivery date using the given formula:

   - **time_to_deliver = order_delivered_customer_date - order_purchase_timestamp**

   - **diff_estimated_delivery = order_delivered_customer_date - order_estimated_delivery_date**

2. **Find out the top 5 states with the highest & lowest average freight value.**

3. **Find out the top 5 states with the highest & lowest average delivery time.**

4. **Find out the top 5 states where the order delivery is really fast as compared to the estimated date of delivery.**
   You can use the difference between the averages of actual & estimated delivery date to figure out how fast the delivery was for each state.

6. **Analysis based on the payments:**

   1. **Find the month on month no. of orders placed using different payment types.**

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

Evaluation Criteria (100 points):

1. **Initial exploration like checking the structure & characteristics of the data (15 points)**

2. **In-depth Exploration (15 points)**

3. **Evolution of E-commerce orders in the Brazil region (10 points)**

4. **Impact on Economy (20 points)**

5. **Analysis on sales, freight and delivery time (20 points)**

6. **Analysis based on the payments (10 points)**

7. **Actionable Insights & Recommendations (10 points)**

Submission Process <IMP>:

Once you're done with the case study...

- **Use a Word document to paste your SQL queries along with a screenshot of the first 10 rows from the output.**
- **List down any valuable insights that you find during the analysis and provide some action items from the company's perspective in order to improve the current situation.**
- **Convert your solutions doc into a PDF, and upload the same on the platform.**
- **Please note that after submitting once, you will not be allowed to edit your submission.**

General Guidelines:

- **Evaluation will be kept lenient, so make sure you attempt this case study.**
- **Try to attempt this before it is discussed in the Live Case Discussion with the Instructor.**
- **It is understandable that you might struggle with getting started on this or feel stuck at some point.**
  In such case:
    - **Read the question carefully and try to understand what exactly is being asked.**
    - **Brainstorm a little. If you're getting an error, remember that Google is your best friend.**
    - **You can watch the lecture recordings or go through your lecture notes once again if you feel like you're getting confused over some specific topics.**
    - **Discuss your problems with your peers. Make use of the Slack channel and WhatsApp group.**
    - **Only if you think that there's a major issue, you can reach out to your Instructor via Slack or Email.**

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

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

   The columns present in the customers table are customer_id,customer_unique_id,

   Customer_zip_code,customer_city,customer_state and the data types of those coloumns are as mentioned below

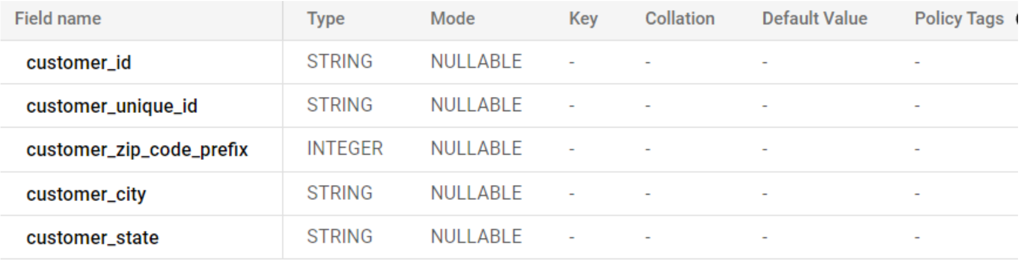
   customer_id=>String

   customer_unique_id=>String

   Customer_zip_code=>Integer

   customer_city=> String

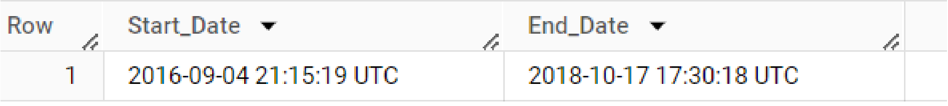
   customer_state=>String



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

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

```
select min(order_purchase_timestamp) as
Start_Date,max(order_purchase_timestamp) as End_Date
from `company.orders`;
```

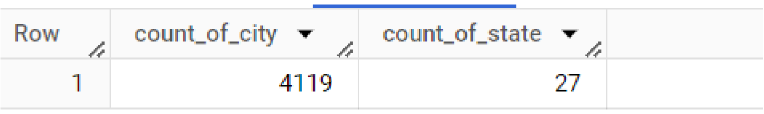


Row	Start_Date	End_Date
1	2016-09-04 21:15:19 UTC	2018-10-17 17:30:18 UTC

As seen above the start date and time for the order when it was placed for the first time is 2016-0-04 at 21:15:19 and the last order which was placed Is on 2018-10-17 at 17:30:18

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

```
select count(distinct a.customer_city) as count_of_city,count(distinct
a.customer_state) as count_of_state
from `company.customers` a inner join
`company.orders` b
on a.customer_id = b.customer_id;
```



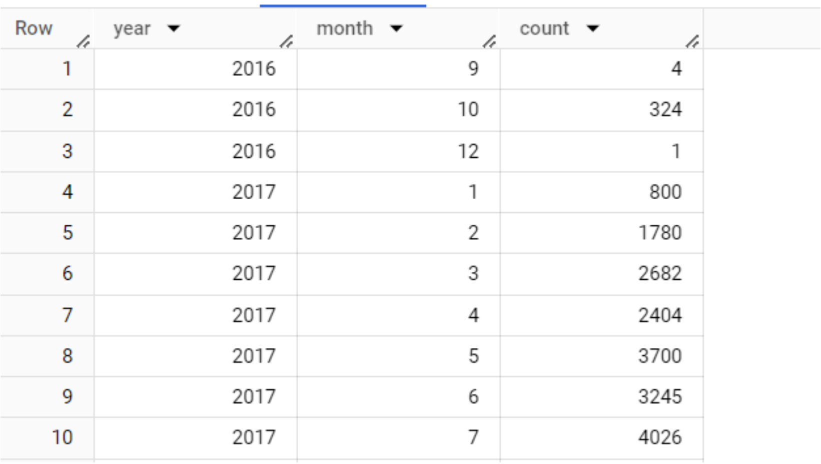
Row	count_of_city	count_of_state
1	4119	27

There are around 4119 different cities and 27 different states from where the customers orders during the time range mentioned in the 1.2.

2. In-depth Exploration:

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

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



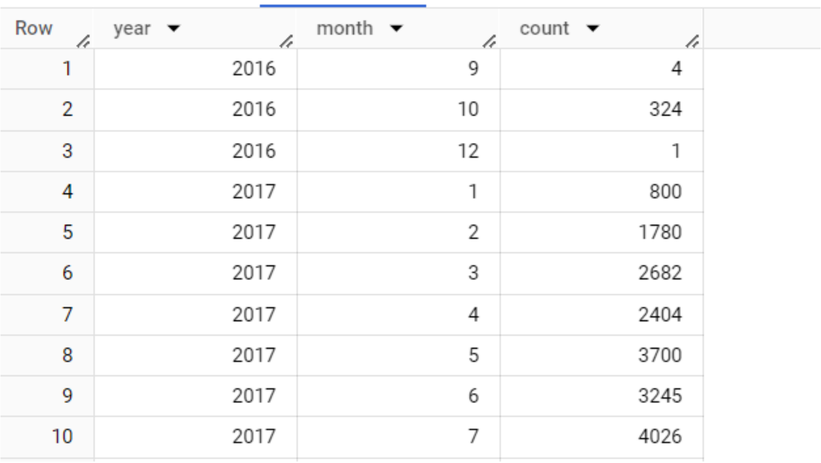
Row	year	month	count
1	2016	9	4
2	2016	10	324
3	2016	12	1
4	2017	1	800
5	2017	2	1780
6	2017	3	2682
7	2017	4	2404
8	2017	5	3700
9	2017	6	3245
10	2017	7	4026

The count of the orders at the beginning was 4 as of 2016 November as seen above post which it was constantly growing till November of 2017 where the count of orders peaked out at 7544 and post which it has been decreasing. However, there was a rise once again in 2018 march at 7211 post which it decreased once again and the latest count of orders which was sold stands at 4 as of 2018 of oct which is same as the starting point of the sales. Most profitable range can be seen in the month of November of 2017 where the orders shot up from 4631 to 7211 in the span of 1 month.

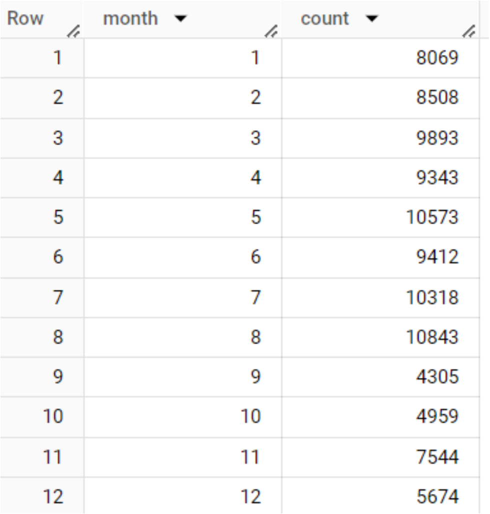
2. Can we see some kind of monthly seasonality in terms of the no. of orders being placed?

```
Q1 select extract(year from order_purchase_timestamp) as year,
extract(month from order_purchase_timestamp) as month,
count(distinct order_id) as count
from `company.orders`
group by 1,2
order by 1,2;

Q2
select
extract(month from order_purchase_timestamp) as month,
count(distinct order_id) as count
from `company.orders`
```

```
group by 1
order by 1;
```



Row	year	month	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

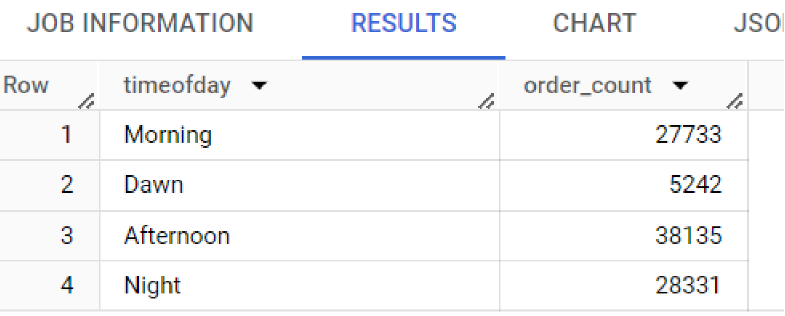


Row	month	count
1	1	8069
2	2	8508
3	3	9893
4	4	9343
5	5	10573
6	6	9412
7	7	10318
8	8	10843
9	9	4305
10	10	4959
11	11	7544
12	12	5674

If we can see the orders which are sold across the month and year from 2016 to 2018. In the month of may there are lot of sales which has taken place That is both in 2017 and 2018 which is 3700,6873 which is in total 10573 across all the three years. This is the month of Mother's Day and upcoming month of the most famous day in brazil that is the Dia dos Namorados which is Brazilian Valentine's Day where it is observed that there are lot of sales that takes place during this month.
Consistency in sales have been observed taking place over the years in the month of July and August which sums up to about 10318 and 10843 across all the three years. This is also the month of Father's day in brazil.
High no of sales is seen in the month of November as these mark the beginning of holiday season and Christmas in Brazil where the sales are very high and consistent topping of 7544 in 2017 and total sales made in this month is 7544

3. During what time of the day, do the Brazilian customers mostly place their orders? (Dawn, Morning, Afternoon or Night)
    1. 0-6 hrs : Dawn
    2. 7-12 hrs : Mornings
    3. 13-18 hrs : Afternoon
    4. 19-23 hrs : Night

```
select
case
when extract(hour from order_purchase_timestamp) <= 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 timeofday,
count(distinct order_id) as order_count
from `company.orders`
group by
timeofday;
```



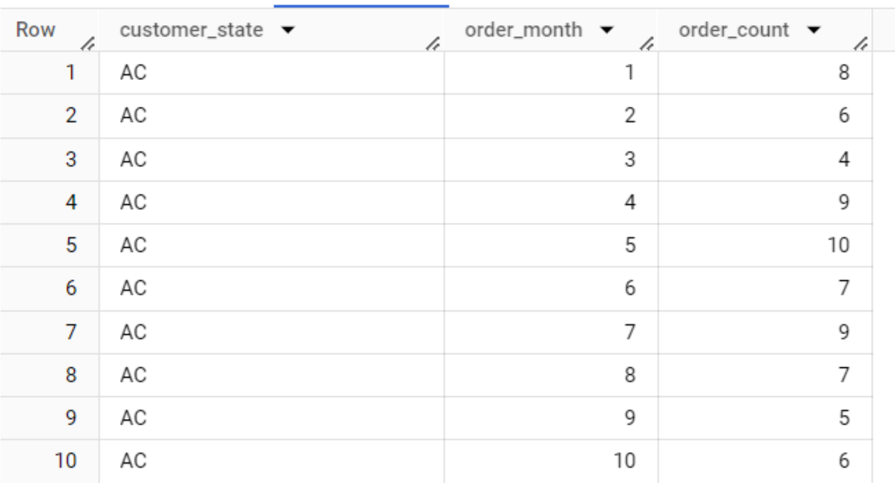
Row	timeofday	order_count
1	Morning	27733
2	Dawn	5242
3	Afternoon	38135
4	Night	28331

We can see from the above table that the highest no of sales was made during the afternoon that is about 38135 and second highest at nigh time.
While dawn is when the sales were the least.

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

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

```
select

b.customer_state,

extract(month from a.order_purchase_timestamp) as order_month,

count(*) as order_count

from `company.orders` a join `company.customers` b

on a.customer_id = b.customer_id

group by 1,2

order by 1,2;
```

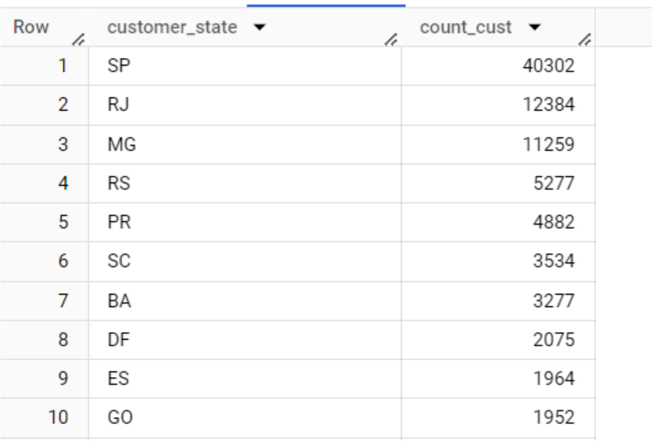


Row	customer_state	order_month	order_count
1	AC	1	8
2	AC	2	6
3	AC	3	4
4	AC	4	9
5	AC	5	10
6	AC	6	7
7	AC	7	9
8	AC	8	7
9	AC	9	5
10	AC	10	6

This is the list of the customer who placed orders month on month across different states on all the three years.

2. How are the customers distributed across all the states?

```
select

customer_state,

count(distinct customer_unique_id) as count_cust

from `company.customers`

group by 1

order by 2 desc;
```



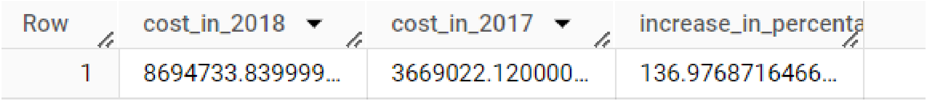
Row	customer_state	count_cust
1	SP	40302
2	RJ	12384
3	MG	11259
4	RS	5277
5	PR	4882
6	SC	3534
7	BA	3277
8	DF	2075
9	ES	1964
10	GO	1952

São Paulo or SP has the highest no of customers and this is also the most populated state of the Brazil and next in line is the Rio de Janeiro which has around 12384 customers post which is the Minas Gerais (MG) which is in third place with the customer count of 11259 there mark the top three states which the highest customer count, while Roraima (RR) has the least no of customer that is around 45.

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

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

```
select sum(case when year = 2018 then payment_value end) as cost_in_2018,

sum(case when year = 2017 then payment_value end) as cost_in_2017,

((sum(case when year = 2018 then payment_value end) - sum(case when year =
2017 then payment_value end))/sum(case when year = 2017 then payment_value
end))*100 as increase_in_percentage

from(

select a.order_id,a.payment_value,

extract(month from b.order_purchase_timestamp) as month,

extract(year from b.order_purchase_timestamp) as year

from `company.payments` a inner join `company.orders` b

on a.order_id = b.order_id

where extract(month from order_purchase_timestamp) between 1 and 8

);
```

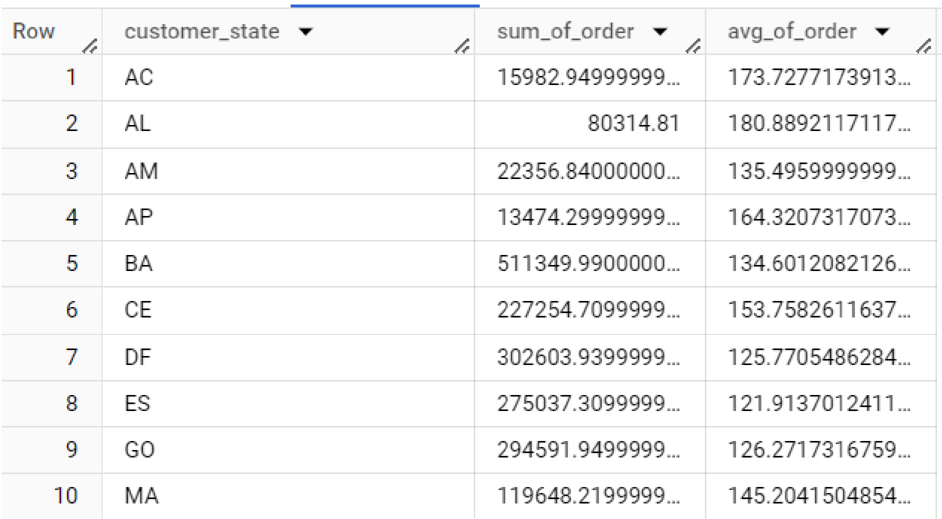


Row	cost_in_2018	cost_in_2017	increase_in_percenta
1	8694733.839999...	3669022.120000...	136.9768716466...

The percentage increase in the cost of the order from 2018 to 2017 in the month from Jan to Aug is 136.97687164665447%

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

```
select customer_state,sum(price) as sum_of_order,avg(price) as avg_of_order

from(

  select a.order_id,b.customer_state,c.price

from `company.orders` a join `company.customers`b

on a.customer_id = b.customer_id

join  `company.order_items` c

on a.order_id = c.order_id
)
group by 1
order by 1;
```

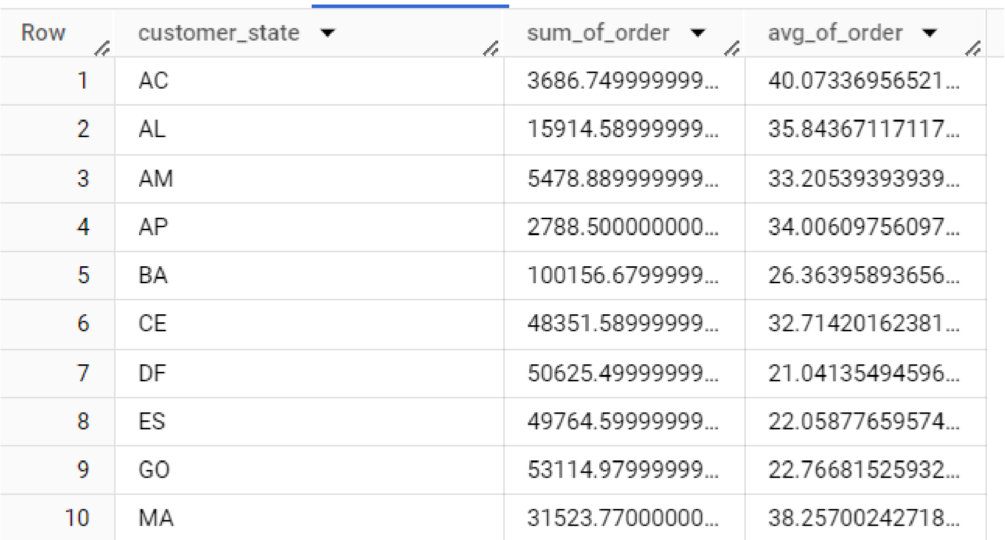


Row	customer_state	sum_of_order	avg_of_order
1	AC	15982.94999999...	173.7277173913...
2	AL	80314.81	180.8892117117...
3	AM	22356.84000000...	135.4959999999...
4	AP	13474.29999999...	164.3207317073...
5	BA	511349.9900000...	134.6012082126...
6	CE	227254.7099999...	153.7582611637...
7	DF	302603.9399999...	125.7705486284...
8	ES	275037.3099999...	121.9137012411...
9	GO	294591.9499999...	126.2717316759...
10	MA	119648.2199999...	145.2041504854...

São Paulo (SP) has the highest order price which is about 5202955.05 and Paraíba (PB) has the high average of sales while Roraima (RR) has the least order price of about 7829.429 and São Paulo (SP) has the least average order price which is about 109.65 again this depends on the no of customers who placed order in that particular state.

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

```
select customer_state,sum(freight_value) as sum_of_order,avg(freight_value) as avg_of_order
from(
  select a.order_id,b.customer_state,c.freight_value
from `company.orders` a join `company.customers`b
on a.customer_id = b.customer_id
join  `company.order_items` c
on a.order_id = c.order_id
)
group by 1
order by 1;
```



Row	customer_state	sum_of_order	avg_of_order
1	AC	3686.749999999...	40.07336956521...
2	AL	15914.58999999...	35.84367117117...
3	AM	5478.889999999...	33.20539393939...
4	AP	2788.500000000...	34.00609756097...
5	BA	100156.6799999...	26.36395893656...
6	CE	48351.58999999...	32.71420162381...
7	DF	50625.49999999...	21.04135494596...
8	ES	49764.59999999...	22.05877659574...
9	GO	53114.97999999...	22.76681525932...
10	MA	31523.77000000...	38.25700242718...

São Paulo (SP) has the highest order freight value 718723.069 and least avg freight value of 15.147 while Roraima (RR) has the highest average value of about 42.98442 and least value of freight that is 2235.19.

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

1. Find the no. of days taken to deliver each order from the order's purchase date as delivery time.
   Also, calculate the difference (in days) between the estimated & actual delivery date of an order.
   Do this in a single query.

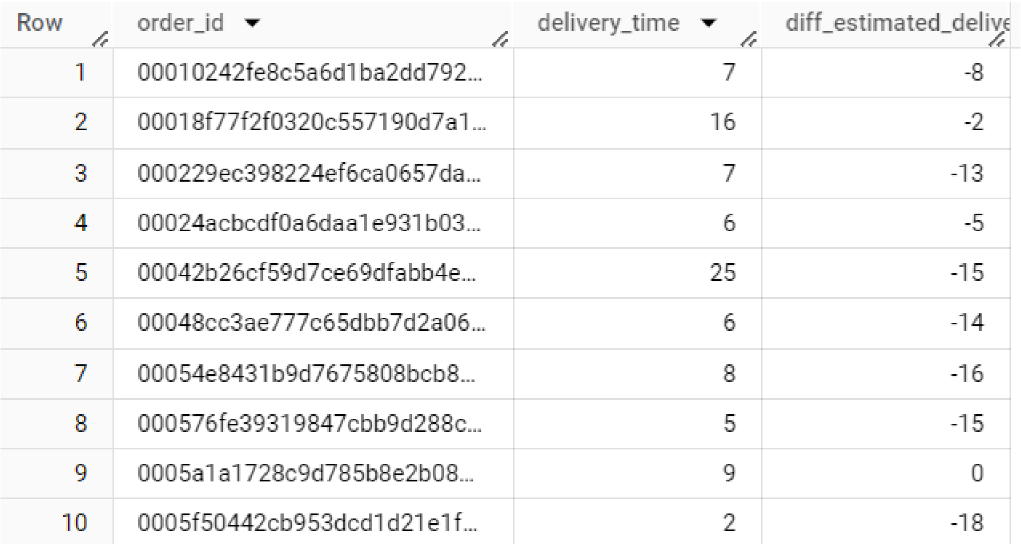
   You can calculate the delivery time and the difference between the estimated & actual delivery date using the given formula:

   ○ **time_to_deliver** = order_delivered_customer_date - order_purchase_timestamp

   ○ **diff_estimated_delivery** = order_delivered_customer_date - order_estimated_delivery_date

```
select order_id,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 `company.orders`

where date_diff(order_delivered_customer_date, order_purchase_timestamp,
day) is not null and
```

```
date_diff(order_delivered_customer_date, order_estimated_delivery_date, day)
is not null

order by 1;
```

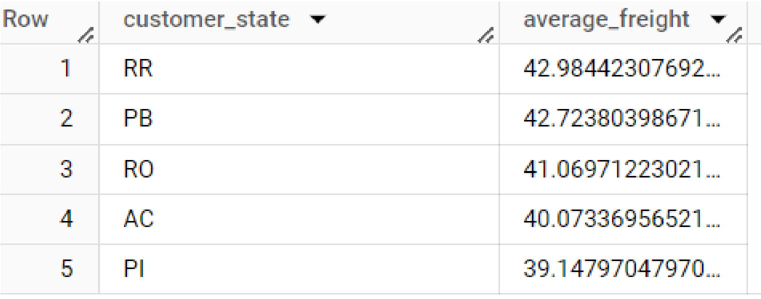


Row	order_id	delivery_time	diff_estimated_delive
1	00010242fe8c5a6d1ba2dd792...	7	-8
2	00018f77f2f0320c557190d7a1...	16	-2
3	000229ec398224ef6ca0657da...	7	-13
4	00024acbcdf0a6daa1e931b03...	6	-5
5	00042b26cf59d7ce69dfabb4e...	25	-15
6	00048cc3ae777c65dbb7d2a06...	6	-14
7	00054e8431b9d7675808bcb8...	8	-16
8	000576fe39319847cbb9d288c...	5	-15
9	0005a1a1728c9d785b8e2b08...	9	0
10	0005f50442cb953dcd1d21e1f...	2	-18

The orders id with null as the delivery time and difference with the estimated delivery means the ones which are not yet delivered or the estimate time is not present and hence the data is filtered out with not null value. If we see the data the maximum time taken for delivery from the purchase date is 209 days for the order id ca07593549f1816d26a572e06dc1eab6 and max estimated time is for the order 1b3190b2dfa9d789e1f14c05b647a14a which is 188 days. The least time taken for delivery of the product which is delivery is e65f1eeee1f52024ad1dcd03447f7482 which was delivered on the same day and least expected estimated delivery time is for the order 0607f0efea4b566f1eb8f7d3c2397320 which is -146 days which goes to say that the order was delivery 146 days before it was expected/estimated.

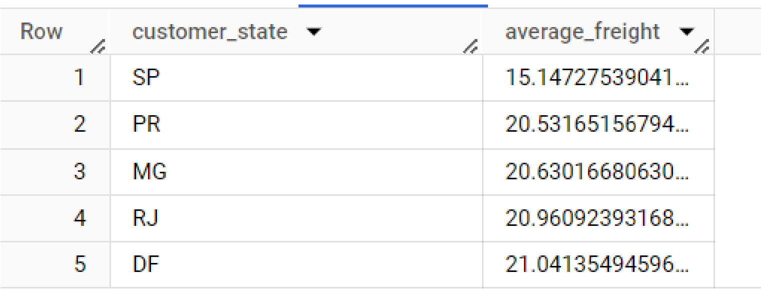
2. Find out the top 5 states with the highest & lowest average freight value.

```
with avg_desc as (
  select customer_state,average_freight
from (
select *,dense_rank() over (order by average_freight desc) as r_no
from (
select a.customer_state,avg(c.freight_value) as average_freight,
from `company.customers` a join `company.orders` b
on a.customer_id = b.customer_id
join `company.order_items` c
on b.order_id = c.order_id
group by 1
) t
) t1
where r_no <= 5
order by average_freight desc),
avg_asc as (
select customer_state,average_freight
from (
select *,dense_rank() over (order by average_freight asc) as r_no
from (
select a.customer_state,avg(c.freight_value) as average_freight,
from `company.customers` a join `company.orders` b
on a.customer_id = b.customer_id
join `company.order_items` c
on b.order_id = c.order_id
group by 1
) t
) t1
where r_no <= 5
order by average_freight asc
)

select *
from avg_desc;
```



Row	customer_state	average_freight
1	RR	42.98442307692...
2	PB	42.72380398671...
3	RO	41.06971223021...
4	AC	40.07336956521...
5	PI	39.14797047970...

```
select *

from avg_asc;
```

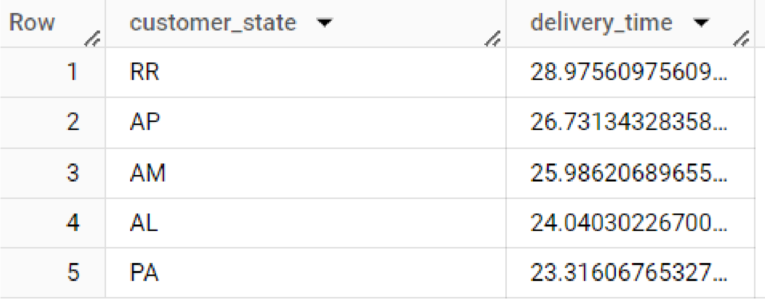


Row	customer_state	average_freight
1	SP	15.14727539041...
2	PR	20.53165156794...
3	MG	20.63016680630...
4	RJ	20.96092393168...
5	DF	21.04135494596...

The first 5 states are the top 5 states with the highest average value and the last 5 states are the top 5 ones with the lowest average values.

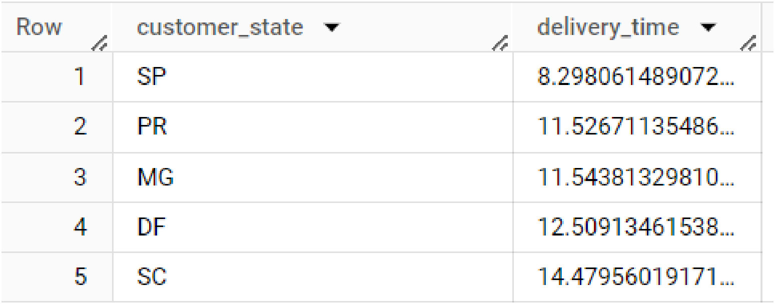
3. Find out the top 5 states with the highest & lowest average delivery time.

```
with avg_desc as (

  select customer_state,delivery_time

from (

select *,dense_rank() over (order by delivery_time desc) as r_no

from (

select a.customer_state,avg(date_diff(b.order_delivered_customer_date,
b.order_purchase_timestamp, day))  as delivery_time

from `company.customers` a join `company.orders` b

on a.customer_id = b.customer_id

group by 1

) t

) t1
where r_no <= 5
order by delivery_time desc),
avg_asc as (
select customer_state,delivery_time
from (
select *,dense_rank() over (order by delivery_time asc) as r_no
from (
select a.customer_state,avg(date_diff(b.order_delivered_customer_date, b.order_purchase_timestamp, day))  as delivery_time
from `company.customers` a join `company.orders` b
on a.customer_id = b.customer_id
group by 1
) t
) t1
where r_no <= 5
order by delivery_time asc
)

select *
from avg_desc;
```



Row	customer_state	delivery_time
1	RR	28.97560975609...
2	AP	26.73134328358...
3	AM	25.98620689655...
4	AL	24.04030226700...
5	PA	23.31606765327...

```
select *
from avg_asc;
```

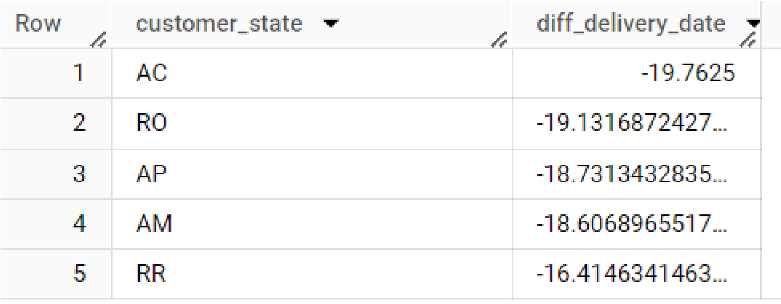


Row	customer_state	delivery_time
1	SP	8.298061489072...
2	PR	11.52671135486...
3	MG	11.54381329810...
4	DF	12.50913461538...
5	SC	14.47956019171...

The first pic represents the top 5 states with the highest delivery time in days and the second pic represent the top 5 states with the lowest delivery time in days.

4. Find out the top 5 states where the order delivery is really fast as compared to the estimated date of delivery.
You can use the difference between the averages of actual & estimated delivery date to figure out how fast the delivery was for each state.

```
select a.customer_state,

avg(date_diff(b.order_delivered_customer_date, b.order_estimated_delivery_date, day)) as diff_delivery_date

from `company.customers` a join `company.orders` b

on a.customer_id = b.customer_id

group by 1

order by 2 asc

limit 5;
```



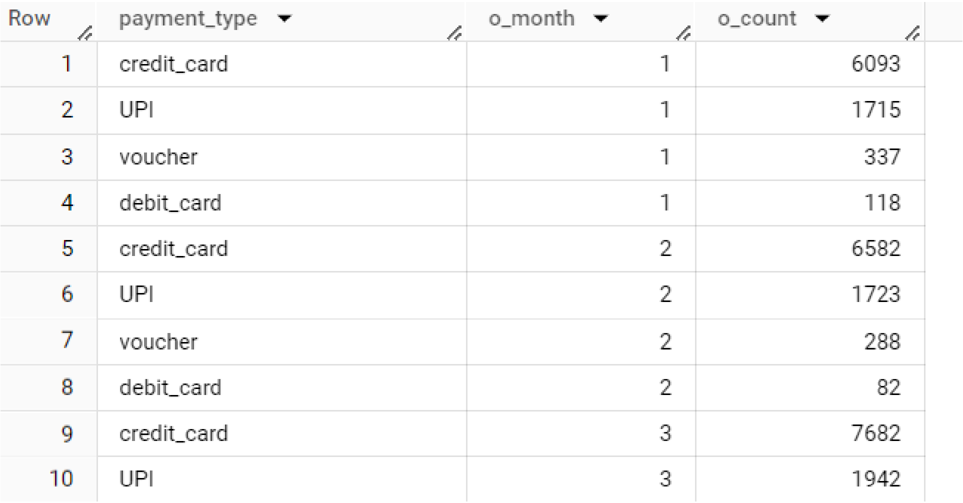
Row	customer_state	diff_delivery_date
1	AC	-19.7625
2	RO	-19.1316872427...
3	AP	-18.7313432835...
4	AM	-18.6068965517...
5	RR	-16.4146341463...

The top 5 states which have had the fastest delivery as show above with Acre(ac) having the fastest delivery rate and Roraima (rr) toping at 5th position on an average.

6. Analysis based on the payments:

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

```
select a.payment_type,
extract(month from b.order_purchase_timestamp) as o_month,count(DISTINCT
b.order_id) as o_count
from `company.payments` a join `company.orders` b
on a.order_id = b.order_id
group by 1,2
order by 2,3 desc;
```



Row	payment_type	o_month	o_count
1	credit_card	1	6093
2	UPI	1	1715
3	voucher	1	337
4	debit_card	1	118
5	credit_card	2	6582
6	UPI	2	1723
7	voucher	2	288
8	debit_card	2	82
9	credit_card	3	7682
10	UPI	3	1942

Based on the data most of the customers placed the orders via credit card type in all the months and the least was placed with the debit card and the upi stands at second of the list in every month as the mode of payment. This is for all the months across all the three years from 2016 to 2018 combined.

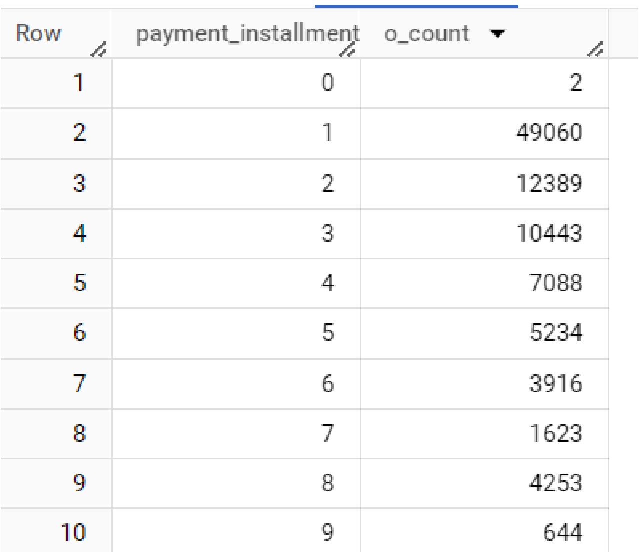
2. Find the no. of orders placed on the basis of the payment installments that have been paid.

select payment_installments,count(distinct order_id) as o_count

from `company.payments`

group by 1

order by 1;



Row	payment_installment	o_count
1	0	2
2	1	49060
3	2	12389
4	3	10443
5	4	7088
6	5	5234
7	6	3916
8	7	1623
9	8	4253
10	9	644

Highest no of orders is placed with the 1 placement installation as seen in the data which is about 49060 and the least were with the 22nd and 23rd instalments where only 1 order was placed.

7. Actionable Insights & Recommendations:

Sales – sales are high during the festive and holiday season of the year, that is during the

The month of May,July,August and November and good discounts and offers could double those values from the current

Time of the day – Most of the sales are seen during the afternoon and night and offers and coupons released during that time would add more value into it

State – it is seen that the most amount of time taken to delivery the product is in Roraima due to which the customers satisfaction will be low and also the sale count in that state is very low. In order to increase it in that area the company has to improve the service by delivering it quicker.

Sao pao being the heavily crowded state of Brazil and the most profitable place for Brazil, Company can increase the sales in that region.

Even though there is a significant increase in the sales from 2017 to 2018 it would even rise if the above mention counter measures are taken by the company.