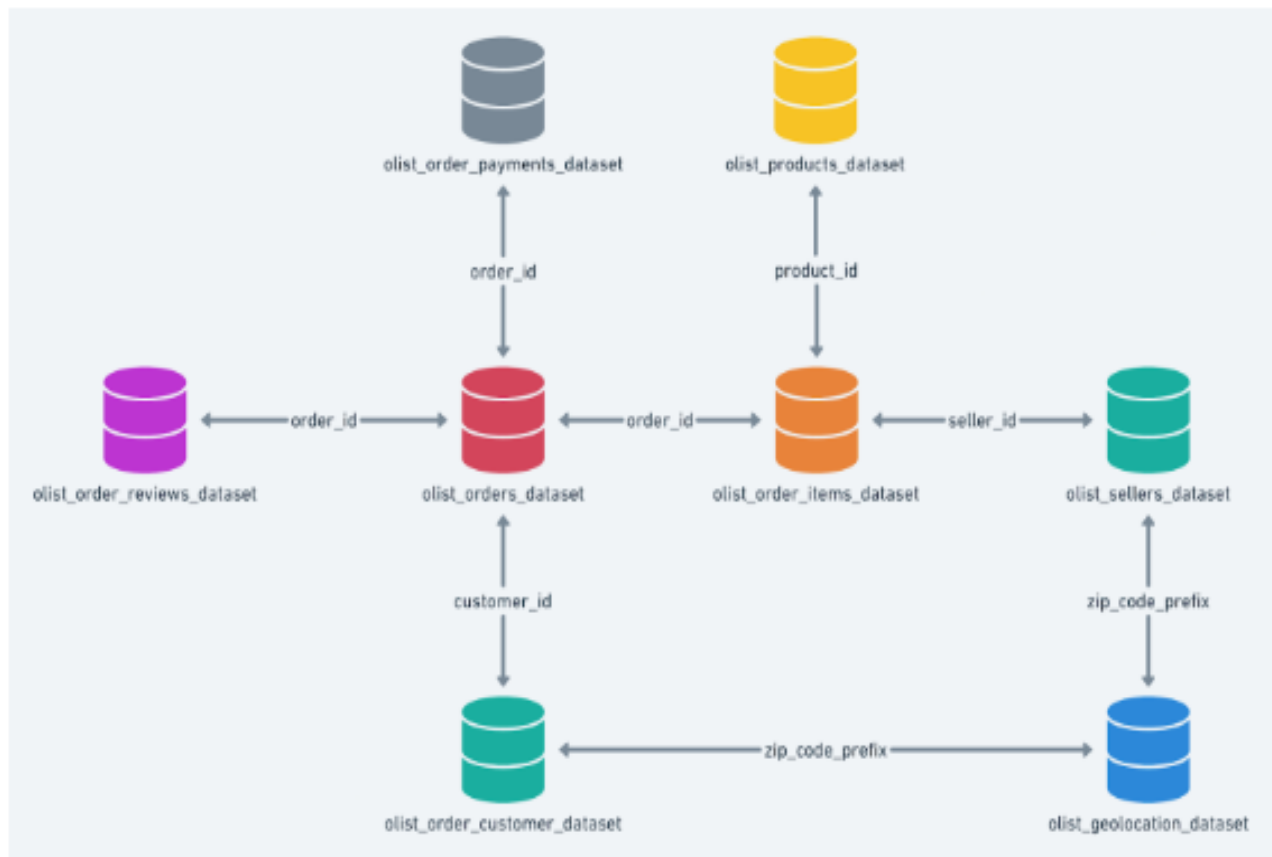


# Business case study about the Brazilian e-commerce

This business case is one of the world's most recognized brands and one of America's leading retailers. It makes itself a preferred shopping destination by offering outstanding value, inspiration, innovation and an exceptional guest experience that no other retailer can deliver. The business case has information of 100k orders from 2016 to 2018 made at Target in Brazil. Its features allows viewing an order from multiple dimensions: from order status, price, payment and freight performance to customer location, product attributes and finally reviews written by customers.





## High level overview of relationship between datasets:



# Business case study about the Brazilian e-commerce

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

1. Data type of columns in a table :
  - a. customer table : String and Integer
  - b. Geolocation table : String,Float and Integer
  - c. Order\_items table : String,Float,Timestamp and Integer
  - d. Order\_reviews table : String,Timestamp and Integer
  - e. Orders table : String and Timestamp
  - f. Payments table : String, Float and Integer
  - g. Products table : String and Integer
  - h. Sellers table : String and Integer

	order_reviews	 QUERY ▾	 SHARE	 COPY
<div>SCHEMADETAILSPREVIEWLINEAGE</div>				
<input type="checkbox"/>	Field name	Type	Mode	Co
<input type="checkbox"/>	<a href="#">review_id</a>	STRING	NULLABLE	
<input type="checkbox"/>	<a href="#">order_id</a>	STRING	NULLABLE	
<input type="checkbox"/>	<a href="#">review_score</a>	INTEGER	NULLABLE	
<input type="checkbox"/>	<a href="#">review_comment_title</a>	STRING	NULLABLE	
<input type="checkbox"/>	<a href="#">review_creation_date</a>	TIMESTAMP	NULLABLE	
<input type="checkbox"/>	<a href="#">review_answer_timestamp</a>	TIMESTAMP	NULLABLE	

# Business case study about the Brazilian e-commerce

2. Time period for which the data is given :

Query :

```
select min(order_purchase_timestamp) as Data_from, max(order_purchase_timestamp) as  
Data_to from `Target.orders`
```

Output:

Untitled 3

RUN

SAVE

SHARE

SCHEDULE

MORE

1

2

3 select min(order\_purchase\_timestamp) as Data\_from, max(order\_purchase\_timestamp) as

4 Data\_to from `Target.orders`

5

Press Alt+F1 for Accessibility C

Query results

SAVE RESULTS

EXPLORE DATA

<

JOB INFORMATION

RESULTS

JSON

EXECUTION DETAILS

EXECUTION GRAP

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

# Business case study about the Brazilian e-commerce

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

Query :

```
select distinct(customer_city) as Cities , customer_state
from `Target.customers`
limit 10
```

Output:

1	Cities	customer_state	
2	acu	RN	
3	ico	CE	
4	ipe	RS	
5	ipu	CE	
6	ita	SC	
7	itu	SP	
8	jau	SP	
9	luz	MG	
10	poa	SP	
11	uba	MG	

# Business case study about the Brazilian e-commerce

## Question 2 : In-depth Exploration:

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

```
SELECT month, count(order_id) as order_count from ( select order_id,  
  
    extract (month from o.order_delivered_customer_date) as Month,  
  
    o.customer_id,  
  
    from `Target.orders` o  
  
    join Target.customers c  
  
    on o.customer_id = c.customer_id)  
  
group by month  
  
having month is not null  
  
order by month
```

	A	B
1	month	order_count
2	1	6880
3	2	7201
4	3	9207
5	4	9699
6	5	10862
7	6	10052
8	7	9294
9	8	12616
10	9	4021
11	10	4705
12	11	4730
13	12	7209

# Business case study about the Brazilian e-commerce

Seasonality with peaks at specific months?

```
SELECT count(order_id) as order_volume,
```

```
EXTRACT (month FROM order_purchase_timestamp) AS month from Target.orders  
group by
```

month

1	order_vol	month
2	7544	11
3	5674	12
4	8508	2
5	9343	4
6	10318	7
7	10573	5
8	4959	10
9	8069	1
10	9412	6
11	4305	9
12	9893	3
13	10843	8

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

Query :

```
select time_hours,
```

# Business case study about the Brazilian e-commerce

```
count(order_id) order_volume
from
(
  select
  order_id,
  case
  when purchase_hours >= 0 and purchase_hours < 6 then 'dawn'
  when purchase_hours >= 6 and purchase_hours < 12 then 'morning'
  when purchase_hours >= 12 and purchase_hours < 18 then 'afternoon'
  when purchase_hours >= 18 and purchase_hours <= 23 then 'night'
  end as time_hours
  from (
    select
    order_id,
    EXTRACT(hour from order_purchase_timestamp) as purchase_hours
    from `Target.orders`
  )
)
group by time_hours
order by order_volume
```

Output:

<		JOB INFORMATION	RESULTS
Row	timehours	order_volume	
1	dawn	4740	
2	morning	22240	
3	night	34100	
4	afternoon	38361	

Hence, from the above output Brazilian customers tend to buy in the afternoon.

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

# Business case study about the Brazilian e-commerce

1. Get month on month orders by states :

Query :

```
SELECT month, customer_state, count(order_id) as order_count
from (
  select order_id,
  extract (month from o.order_delivered_customer_date) as Month,
  o.customer_id,
  customer_state
  from `Target.orders` o
  join Target.customers c
  on o.customer_id = c.customer_id)
group by customer_state, month
order by month
limit 10
```

Output:

	A	B	C		A	B	C
1	month	customer_state	order_count	1	month	customer_state	order_count
2	1	CE	100	2	2	CE	57
3	1	RJ	931	3	2	SC	291
4	1	MT	66	4	2	PR	396
5	1	RS	352	5	2	PA	58
6	1	PR	369	6	2	MG	897
7	1	SP	2731	7	2	SP	3046
8	1	BA	247	8	2	BA	213
9	1	SC	249	9	2	RJ	942
10	1	MG	815	10	2	MA	47
11	1	ES	145	11	2	RS	371

## 2.Distribution of customers across the states in Brazil



# Business case study about the Brazilian e-commerce

Query :

```
SELECT customer_state, COUNT(DISTINCT customer_id) AS num_customers
FROM Target.customers
GROUP BY customer_state
limit 10
```

Output:

	A	B	
1	customer_state	num_customers	
2	RN	485	
3	CE	1336	
4	RS	5466	
5	SC	3637	
6	SP	41746	
7	MG	11635	
8	BA	3380	
9	RJ	12852	
10	GO	2020	
11	MA	747	
12			

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

# Business case study about the Brazilian e-commerce

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

Query :

```
SELECT year, month, total_sales,
round((total_sales-LAG(total_sales)over(order by month))/LAG(total_sales)over (order
by month)*100,2)percent_change_sales
from (
  select
  year, month, round(sum(payment_value),0)total_sales
  from (
    select
    extract (year from order_purchase_timestamp)as year,
    extract(month from order_purchase_timestamp)as month,
    o.order_id,
    payment_value
    from Target.orders o join Target.payments p
    on o.order_id = p.order_id
  )table_one
  group by year, month
  order by month
)
where year in (2017, 2018) and month between 1 and 8
order by year
```

Output:

	A	B	C	D		A	B	C	D	E
1	year	month	total_sales	percent_change_sales	1	year	month	total_sale	percent_change_sales	
2	2017	6	511276	-13.77	2	2018	4	1160785	0.1	
3	2017	4	417788	-7.13	3	2018	5	1153982	-0.59	
4	2017	1	138488		4	2018	3	1159652	16.85	
5	2017	7	592383	15.86	5	2018	7	1066541	4.17	
6	2017	3	449864	54.11	6	2018	1	1115004		
7	2017	8	674396	13.84	7	2018	2	992463	-10.99	
8	2017	5	592919	41.92	8	2018	8	1022425	-4.14	
9	2017	2	291908	110.78	9	2018	6	1023880	-11.27	
10					10					

## 2. Mean & Sum of price and freight value by customer state

Query :

```
select c.customer_state,
```

# Business case study about the Brazilian e-commerce

```
round (sum(i.price),1) sumcost_bystate,  
round (sum(i.freight_value),1) fright_value_bystate,  
round(avg(i.price),1) mean_cost,  
round(avg(i.freight_value),1) mean_frieght_c  
from Target.orders o join `Target.order_items` i  
on o.order_id = i.order_id  
join Target.customers c  
on o.customer_id = c.customer_id  
group by c.customer_state  
order by mean_cost desc, mean_frieght_c desc
```

Output:

	A	B	C	D	E
1	customer_state	sumcost_bystate	fright_value_bystate	mean_cost	mean_frieght_c
2	PB	115268.1	25719.7	191.5	42.7
3	AL	80314.8	15914.6	180.9	35.8
4	AC	15982.9	3686.7	173.7	40.1
5	RO	46140.6	11417.4	166	41.1
6	PA	178947.8	38699.3	165.7	35.8
7	AP	13474.3	2788.5	164.3	34
8	PI	86914.1	21218.2	160.4	39.1
9	TO	49621.7	11732.7	157.5	37.2
10	RN	83035	18860.1	157	35.7
11	CE	227254.7	48351.6	153.8	32.7
12	SE	58920.9	14111.5	153	36.7

## Question 5. Analysis on sales, freight and delivery time

1. Calculate days between purchasing, delivering and estimated delivery

Query :

# Business case study about the Brazilian e-commerce

```
select order_id,  
date_diff(order_delivered_customer_date, order_purchase_timestamp, day) delivery,  
date_diff(order_estimated_delivery_date, order_purchase_timestamp, day) estimated,  
date_diff(order_estimated_delivery_date, order_delivered_carrier_date, day)  
days_estimated  
from `Target.orders`  
where order_delivered_customer_date is not null  
order by delivery, estimated
```

Output:

	A	B	C	D
1	order_id	delivery	estimated	days_estimated
2	d5fbeedc85190ba88	0	8	8
3	79e324907160caea5	0	9	9
4	e65f1eeee1f52024ac	0	10	10
5	1d893dd7ca5f77ebf5	0	10	10
6	b70a8d75313560b4a	0	10	10
7	d3ca7b82c922817b0	0	12	12
8	f3c6775ba3d2d9fe28	0	12	12
9	21a8ffca665bc7a108	0	12	12
10	f349cdb62f69c3fae5c	0	13	13
11	38c1e3d4ed6a13cd0	0	17	16
12	434cecee7d1a65fc65	0	20	20

2. Find time\_to\_delivery & diff\_estimated\_delivery. Formula for the same given below:

- time\_to\_delivery =  
order\_purchase\_timestamp-order\_delivered\_customer\_date
- diff\_estimated\_delivery =  
order\_estimated\_delivery\_date-order\_delivered\_customer\_date

# Business case study about the Brazilian e-commerce

Query :

```
select
date_diff(order_purchase_timestamp,order_delivered_customer_date, day)
time_to_delivery,
date_diff(order_estimated_delivery_date, order_delivered_customer_date, day)
diff_estimated_delivery
from `Target.orders`
where order_purchase_timestamp is not null and order_estimated_delivery_date is not
null
```

Output:

	A	B	C	D
1	time_to_delivery	diff_estimated_delivery		
2	-30	-12		
3	-30	28		
4	-35	16		
5	-30	1		
6	-32	0		
7	-29	1		
8	-43	-4		
9	-40	-4		
10	-37	-1		
11	-33	-5		
12	-38	-6		

# Business case study about the Brazilian e-commerce

3. Group data by state, take mean of freight\_value, time\_to\_delivery, diff\_estimated\_delivery

Query :

```
select c.customer_state,

round(avg(i.freight_value),2) mean_,

round(avg(date_diff(o.order_delivered_customer_date, o.order_purchase_timestamp,
day)),2)as delivery_days,

round(avg(date_diff(o.order_estimated_delivery_date,
o.order_delivered_carrier_date,day)),2)as estimated

from Target.orders o join `Target.customers` c

on o.customer_id = c.customer_id

join Target.order_items i

on o.order_id = i.order_id

group by c.customer_state
```

# Business case study about the Brazilian e-commerce

	A	B	C	D
1	customer_	mean_	delivery_days	estimated
2	MT	28.17	17.51	28.29
3	MA	38.26	21.2	26.61
4	AL	35.84	23.99	28.49
5	SP	15.15	8.26	15.68
6	MG	20.63	11.52	21
7	PE	32.92	17.79	27.46
8	RJ	20.96	14.69	22.72
9	DF	21.04	12.5	20.88
10	RS	21.74	14.71	24.98
11	SE	36.65	20.98	26.62

4. Sort the data to get the following:

5. Top 5 states with highest/lowest average freight value - sort in desc/asc limit 5

Query :

```
select c.customer_state,round(avg(i.freight_value),2) mean_,  
  
round(avg(date_diff(o.order_delivered_customer_date, o.order_purchase_timestamp,  
day)),2)as delivery_days,  
  
round(avg(date_diff(o.order_estimated_delivery_date,  
o.order_delivered_carrier_date,day)),2)as estimated  
  
from Target.orders o join `Target.customers` c on o.customer_id = c.customer_id  
  
join Target.order_items i on o.order_id = i.order_id  
  
group by c.customer_state order by mean_ asc
```

## Business case study about the Brazilian e-commerce

	A	B	C	D
1	customer_	mean_	delivery_d	destimated
2	SP	15.15	8.26	15.68
3	PR	20.53	11.48	21.06
4	MG	20.63	11.52	21
5	RJ	20.96	14.69	22.72
6	DF	21.04	12.5	20.88
7				

Desc

	A	B	C	D
1	customer_	mean_	delivery_d	destimated
2	RR	42.98	27.83	40.75
3	PB	42.72	20.12	28.92
4	RO	41.07	19.28	35.86
5	AC	40.07	20.33	37.3
6	PI	39.15	18.93	26.67
7				

### 6. Top 5 states with highest/lowest average time to delivery

Lowest average time to deliver :

	A	B	C	D
1	customer_	mean_	delivery_d	destimated
2	SP	15.15	8.26	15.68
3	PR	20.53	11.48	21.06
4	MG	20.63	11.52	21
5	DF	21.04	12.5	20.88
6	SC	21.47	14.52	22.1
7				



## Business case study about the Brazilian e-commerce

Highest average time to deliver :

	A	B	C	D	E
1	customer_	mean_	delivery_d	estimated	
2	RR	42.98	27.83	40.75	
3	AP	34.01	27.75	42	
4	AM	33.21	25.96	42.29	
5	AL	35.84	23.99	28.49	
6	PA	35.83	23.3	33.53	
7					
8					

### 7. Top 5 states where delivery is really fast/ not so fast compared to estimated date

States Not so fast to estimated date :

	A	B	C	D	E
1	customer_	mean_	delivery_d	estimated	
2	SP	15.15	8.26	15.68	
3	DF	21.04	12.5	20.88	
4	MG	20.63	11.52	21	
5	PR	20.53	11.48	21.06	
6	ES	22.06	15.19	21.8	
7					

# Business case study about the Brazilian e-commerce

Fastest States to estimated date :

	A	B	C	D	E
1	customer_	mean_	delivery_d	estimated	
2	AM	33.21	25.96	42.29	
3	AP	34.01	27.75	42	
4	RR	42.98	27.83	40.75	
5	AC	40.07	20.33	37.3	
6	RO	41.07	19.28	35.86	
7					

Question 6. Payment type analysis:

1. Month over Month count of orders for different payment types

Query :

```
select payment_type, month, count(order_id) order_c
from (
  select
    p.payment_type,
    extract (month from o.order_purchase_timestamp) month,
    p.order_id
  from Target.payments p join Target.orders o
  on
    p.order_id = o.order_id)
group by payment_type, month
order by payment_type, month
limit 10
```

# Business case study about the Brazilian e-commerce

Output:

	A	B	C
1	payment_	month	order_c
2	UPI	1	1715
3	UPI	2	1723
4	UPI	3	1942
5	UPI	4	1783
6	UPI	5	2035
7	UPI	6	1807
8	UPI	7	2074
9	UPI	8	2077
10	UPI	9	903
11	UPI	10	1056

## 2. Count of orders based on the no. of payment installments

Query :

```
select payment_installments,  
count(order_id) order_volume  
from `Target.payments`  
group by payment_installments
```

# Business case study about the Brazilian e-commerce

	A	B	C
1	payment	order_volume	
2	0	2	
3	1	52546	
4	2	12413	
5	3	10461	
6	4	7098	
7	5	5239	
8	6	3920	
9	7	1626	
10	8	4268	
11	9	644	
12	10	5328	

# Business case study about the Brazilian e-commerce

## Actionable Insights :

### Observations in Detail:

- August had the highest number of orders, while September had the lowest. Order volume showed an increasing trend in Q1 (January to March) and Q4 (October to December).
- The percentage of orders in August was the highest among all months.
- There was an increase in order volume and sales value from 2017 to 2018.
- Brazilians tend to shop more in the afternoon and less during dawn.
- The state of SP had the highest number of orders among all states in Brazil.
- The state of RR had the lowest number of orders among all states in Brazil.
- The state of PB had an average cost per order, while the state of SP had the lowest average cost per order.
- August had the highest number of orders, while September had the lowest. Order volume showed an increasing trend in Q1 (January to March) and Q4 (October to December).
- The percentage of orders in August was the highest among all months.
- There was an increase in order volume and sales value from 2017 to 2018.
- Brazilians tend to shop more in the afternoon and less during dawn.
- The state of SP had the highest number of orders among all states in Brazil.
- The state of RR had the lowest number of orders among all states in Brazil.
- The state of PB had an average cost per order, while the state of SP had the lowest average cost per order.

# **Business case study about the Brazilian e-commerce**

## **Recommendations :**

- **It is recommended to increase delivery speed despite most orders being delivered within five days, as some orders are taking more than 100 days to deliver.**
- **Detailed research is needed to determine the reason for very low sales in certain states, such as RR, AP, AC, and others.**
- **A detailed study is necessary to consider opening new stores in states with high order volumes, such as SP.**
- **Since most payments are made through credit cards, offering benefits to credit card customers such as low interest or no-cost interest could further boost sales.**
- **The estimated delivery time is always longer than the actual delivery time. To provide customers with accurate information regarding their product, the estimated delivery time should be mentioned closer to the actual delivery time.**
- **August had the highest number of orders, so it is important to ensure that store operating times, delivery times, and office staff are adequate to provide customer satisfaction.**