

SQL Target Business Case Study

> Exploratory Analysis

Given is a data set containing the order information from purchases made at Target in Brazil.

Range of Data:

The first available information of the order was placed on 2016-09-04 at 21:15:19 UTC and the last available information of the order was placed on 2018-10-17 at 17:30:18 UTC.

The following query can be used to extract the above information:

```
select min(order_purchase_timestamp) as first_order,  
max(order_purchase_timestamp) as last_order  
from `target_case_study.orders`;
```

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

Geographical Information

During this time, the orders were placed from 8011 different cities across 27 different states.

The following query can be used to extract the above information:

```
select count(distinct geolocation_city) as distinct_cities,  
count(distinct geolocation_state) as distinct_states  
from `target_case_study.geolocation`;
```

Row	distinct_cities	distinct_states
1	8011	27

> In-depth Exploration of the Evolution of e-Commerce Orders in Brazil and its Impact on Economy

Preferred Time of Shopping

From the 'order_purchase_timestamp' column in the orders table, it can be deduced that 22428 Brazilians have ordered during Morning hours (between 5 AM to 12 PM), 32212 Brazilians have ordered during the Afternoons (between 12 PM to 5 PM) and 24094 number of Brazilians have ordered during the Evenings (between 5 PM to 9 PM).

The following 4 queries can be used to find the above information:

Query to find the number of orders placed during the Morning hours (between 5 AM to 12 AM):

```
select count(order_id)
from `target_case_study.orders`
where time(order_purchase_timestamp) between '05:00:00' and
'12:00:00';
```

Query to find the number of orders placed during the Afternoon hours (between 12 AM to 12 AM):

```
select count(order_id)
from `target_case_study.orders`
where time(order_purchase_timestamp) between '12:00:00' and
'17:00:00';
```

Query to find the number of orders placed during the Evening hours (between 5 PM to 12 AM):

```
select count(order_id)
from `target_case_study.orders`
where time(order_purchase_timestamp) between '17:00:00' and
'21:00:00';
```

Query to find the number of orders placed during the Night hours (between 12 AM to 5 AM):

```
select count(order_id)
from `target_case_study.orders`
where time(order_purchase_timestamp) between '21:00:00' and
'05:00:00';
```

Serial #	orders_placed_m ornings 5 AM to 12 PM	orders_placed_aft ernoons 12 PM to 5 PM	orders_placed_ev enings 5 PM to 9 PM	orders_placed_nig hts 9 PM to 5
1	22428	32212	24094	0

It is interesting to note that no orders were placed during the Night hours.

Month-wise Purchase Trend

The month-wise shopping trend can be extracted from the following query:

```
select count(order_id) as orders_month_wise, purchase_month
from (
  select order_id, extract(month from
date(order_purchase_timestamp)) as purchase_month
  from `target_case_study.orders`
)
group by purchase_month
order by purchase_month;
```

Month	Number of Orders
January	8069
February	8508

March	9893
April	9394
May	10573
June	9412
July	10318
August	10843
September	4305
October	4959
November	7544
December	5674

The following deductions can be made from the extracted data shown in the table above:

1. There is an increase in the number of orders from February to March, this is mostly due to "Carnival" which is the most popular festival celebrated in Brazil.
2. An increase in the number of orders is seen in the month of May and this trend continues until the end of August. This is because all the major festivals in Brazil are celebrated during this period. Starting from "Festa Junia" and "Parintins Folklore Festival" in June and mid-June respectively.
3. The highest number of orders were placed in the month of August, which is when the "Festival de Cachaca" is celebrated.
4. There is a decline in the number of orders placed in the months following August, and the least number of orders are placed in December. An analysis can be made that Christmas is not one of the big festivals that are celebrated in Brazil.

The conclusion from all of this is that festivals are a major contributing factor in the purchase trends in Brazil. And the sales are highest during the months of May and June due to "Festa Junia" and "Parintins Folklore Festival", both of which are celebrated during this time.

Note: Goto link 1 in the Reference section, to find more about festivals celebrated in Brazil.

Number of Customers across all Brazilian States

The query to extract the number of customers per state is as follows:

```
select count(distinct customer_id) as number_of_customers,
customer_state
from `target_case_study.customers`
group by customer_state;
```

Serial #	State	Number of Customers
1	RN	485
2	CE	1336

3	RS	5466
4	SC	3637
5	SP	41746
6	MG	11635
7	BA	3380
8	RJ	12852
9	GO	2020
10	MA	747
11	PE	1652
12	PB	536
13	ES	2033
14	PR	5045
15	RO	253
16	MS	715
17	PA	975
18	TO	280
19	MT	907
20	PI	495
21	AL	413
22	AM	148
23	DF	2140
24	SE	350
25	RR	46
26	AP	68
27	AC	81

The conclusion from the above data is that the greatest number of customers that Target has are based out of the SP (Sao Paulo) state, and the least number of customers are based out of the RR (Roraima) state.

Furthermore, to support this trend, the population density map of Brazil can be brought into the picture (below). It can be observed that the maximum population is based in the southeastern part of the country and all the major cities like Sao Paulo, Rio de Janeiro and Belo Horizonte are located in this region. And the least population in Brazil lives in the Amazonian forest, which is where the state of RR (Roraima) is located.

Note: Information on the Brazilian States and their shorthand representation can be found in link 2 in the reference section.
Also refer to image 1 (scroll below to page 13) to find the population density map of Brazil.

> Impact on Economy

Sales Growth 2017 to 2018 for the time from January to August (both included)

The query to extract the cost of orders from 2017 (from January to August only, both included):

```
select round(sum(payment_value), 2) as total_cost_2017
from `target_case_study.orders` as o full outer join
`target_case_study.payments` as p on o.order_id = p.order_id
where date(o.order_purchase_timestamp) between '2017-01-01' and
'2017-08-31';
```

The query to extract the cost of orders from 2018 (from January to August only, both included):

```
select round(sum(payment_value), 2) as total_cost_2018
from `target_case_study.orders` as o full outer join
`target_case_study.payments` as p on o.order_id = p.order_id
where date(o.order_purchase_timestamp) between '2018-01-01' and
'2018-08-31';
```

The total cost of orders in 2017 was 3669022.12

The total cost of orders in 2018 was 8694733.84

$8694733.84 - 3669022.12 = 5025711.72$

$8694733.84 + 3669022.12 = 12363755.96$

From the above two figures, the percentage increase in the cost of orders from 2017 to 2018 can be calculated as follows:

$(5025711.72 / 12363755.96) * 100 = 0.406487 * 100 = 40.65\%$

Therefore, there has been a **40.65%** increase in the cost of orders from 2017 to 2018 for the same duration of time (from January to August, both included).

Pricing and Freight Cost

The query to find the mean and sum of price and freight value by customer state is as follows:

```
select customer_state, round(avg(price), 2) as mean_price,
round(avg(freight_value), 2) as mean_freight_value, round(sum(price),
2) as sum_of_price, round(sum(freight_value), 2) as sum_of_freight
from `target_case_study.orders` as ord left outer join
`target_case_study.customers` as cust on ord.customer_id =
```

```

cust.customer_id left outer join `target_case_study.order_items`
as oi on ord.order_id = oi.order_id
group by customer_state;

```

Serial #	customer_state	mean_price	mean_freight_value	sum_of_price	sum_of_freight
1	RJ	125.12	20.96	1824092.67	305589.31
2	RS	120.34	21.74	750304.02	135522.74
3	SP	109.65	15.15	5202955.05	718723.07
4	DF	125.77	21.04	302603.94	50625.5
5	PR	119.0	20.53	683083.76	117851.68
6	MT	148.3	28.17	156453.53	29715.43
7	MA	145.2	38.26	119648.22	31523.77
8	AL	180.89	35.84	80314.81	15914.59
9	MG	120.75	20.63	1585308.03	270853.46
10	PE	145.51	32.92	262788.03	59449.66
11	SE	153.04	36.65	58920.85	14111.47
12	PA	165.69	35.83	178947.81	38699.3
13	BA	134.6	26.36	511349.99	100156.68
14	CE	153.76	32.71	227254.71	48351.59
15	GO	126.27	22.77	294591.95	53114.98
16	ES	121.91	22.06	275037.31	49764.6
17	SC	124.65	21.47	520553.34	89660.26
18	PI	160.36	39.15	86914.08	21218.2
19	PB	191.48	42.72	115268.08	25719.73
20	RN	156.97	35.65	83034.98	18860.1
21	AM	135.5	33.21	22356.84	5478.89
22	RR	150.57	42.98	7829.43	2235.19
23	MS	142.63	23.37	116812.64	19144.03
24	TO	157.53	37.25	49621.74	11732.68
25	AC	173.73	40.07	15982.95	3686.75
26	RO	165.97	41.07	46140.64	11417.38
27	AP	164.32	34.01	13474.3	2788.5

> Analysis of Sales, Freight and Delivery Time

Use to below query to create a view which includes the required columns from “orders”, “order_items” and “customers” tables:

```
create view target_case_study.ord_oi_cust as (select ord.order_id,
order_purchase_timestamp, order_delivered_customer_date,
order_estimated_delivery_date,
datetime_diff(order_delivered_customer_date,
order_purchase_timestamp, day) as time_to_delivery,
datetime_diff(order_delivered_customer_date,
order_estimated_delivery_date, day) as diff_estimated_delivery,
oi.freight_value, cust.customer_state
from `target_case_study.orders` as ord left outer join
`target_case_study.customers` as cust on ord.customer_id =
cust.customer_id left outer join `target_case_study.order_items`
as oi on ord.order_id = oi.order_id)
```

IMPORTANT: Run the above query to create a View before proceeding further.

The average of freight_value, time_of_delivery and diff_estimated_delivery can be found using the below query:

```
select avg(time_to_delivery), avg(diff_estimated_delivery),
avg(freight_value), customer_state
from `target_case_study.ord_oi_cust`
group by customer_state;
```

Top 5 States with the Highest Freight Value

The top 5 states with the highest average freight value can be found with the query below:

```
select avg(freight_value) as avg_freight_value, customer_state
from `target_case_study.ord_oi_cust`
group by customer_state
order by avg_freight_value desc
limit 5;
```

Rank	State
1	RR
2	PB
3	RO
4	AC
5	PI

The conclusion from the above table is that the remotes and the sparsely populated regions in Brazil have the highest average freight value.

Top 5 States with the Lowest Freight Value

The top 5 states with the lowest average freight value can be found with the query below:

```
select avg(freight_value) as avg_freight_value, customer_state
from `target_case_study.ord_oi_cust`
group by customer_state
order by avg_freight_value asc
limit 5;
```

Rank	State
1	SP
2	PR
3	MG
4	RJ
5	DF

The conclusion from the above table is that the densely populated regions in Brazil have the lowest average freight value.

Top 5 States with the Highest Average Time of Delivery

The top 5 states with the highest average time of delivery can be found with the query below:

```
select round(avg(time_to_delivery), 2) as avg_tod, customer_state
from `target_case_study.ord_oi_cust`
group by customer_state
order by avg_tod desc
limit 5;
```

Rank	State
1	RR
2	AP
3	AM
4	AL
5	PA

The conclusion from the above table is that the remotes and the sparsely populated regions in Brazil have the highest average time of delivery.

Top 5 States with the Lowest Average Time of Delivery

The top 5 states with the lowest average time of delivery can be found with the query below:

```
select round(avg(time_to_delivery), 2) as avg_tod, customer_state
from `target_case_study.ord_oi_cust`
group by customer_state
order by avg_tod asc
limit 5;
```

Rank	State
1	SP
2	PR
3	MG
4	DF
5	SC

The conclusion from the above table is that the densely populated regions in Brazil have the lowest average time of delivery.

Top 5 States with the Fastest Delivery Time

The top 5 states with the fastest delivery time can be found with the query below:

```
select round(avg(diff_estimated_delivery), 2) as fastest_delivery,
customer_state
from `target_case_study.ord_oi_cust`
group by customer_state
order by fastest_delivery desc
limit 5;
```

Rank	State
1	AL
2	MA
3	SE
4	ES
5	BA

The conclusion from the above table is that the states located in the North-Eastern region of Brazil have the fastest delivery time. This is due to the fact that major shipping ports are located in this part of the country.

Top 5 States with the Slowest Delivery Time

The top 5 states with the slowest delivery time can be found with the query below:

```
select round(avg(diff_estimated_delivery), 2) as slowest_delivery,
customer_state
from `target_case_study.ord_oi_cust`
group by customer_state
order by slowest_delivery asc
limit 5;
```

Rank	State
1	AC
2	RO
3	AM
4	AP
5	RR

The conclusion from the above table is the state located in the Northern region of Brazil has the slowest delivery time. This is due to the fact that this region has the Amazon rainforest, and because of the forest cover it is not very well connected with the rest of the country. Hence, the transport of orders becomes quite a task.

Note: Refer to Image 2 to find the Map of Brazil (page 14).

> Payment Type Analysis

Preferred Mode/ Popular Mode of Payment

The query to extract the month-over-month count of orders for different payment types is as follows:

```
select count(o.order_id), p.payment_type, extract(month from
date(o.order_purchase_timestamp)) as purchase_month
from `target_case_study.orders` as o full outer join
`target_case_study.payments` as p on o.order_id = p.order_id
group by purchase_month, p.payment_type
order by p.payment_type, purchase_month;
```

The table below shows the total of all orders for each payment type (the query above outputs the month-wise total of all orders for each payment type).

Serial #	Payment Type	Number of Orders
1	Credit Card	76795
2	UPI	19784
3	Voucher	5481
4	Debit Card	1529
5	Not Defined	3

6	NULL	1
---	------	---

It can be observed that Credit Card seems to be the preferred mode of payment for Brazilians, followed by UPI and then Vouchers.

Preferred Number of Payment Instalments

The query to extract the count of orders based on the number of payment instalments is as follows:

```
select payment_installments, count(order_id) as order_count
from `target_case_study.payments`
group by payment_installments
order by order_count desc;
```

Serial #	Number of Payment Instalments	Order Count
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
11	12	133
12	15	74
13	18	27
14	11	23
15	24	18
16	20	17
17	13	16
18	14	15
19	17	8
20	16	5
21	21	3
22	0	2

23	22	1
24	23	1

It is observed that the payment for the maximum number of orders has been done in 1 instalment. It is also observed that the maximum number of orders have been paid for in 10 instalments at most. Furthermore, it can also be derived that the customers prefer to make the payment in the least number of instalments.

Image. 1

The image below shows the population density across Brazil.

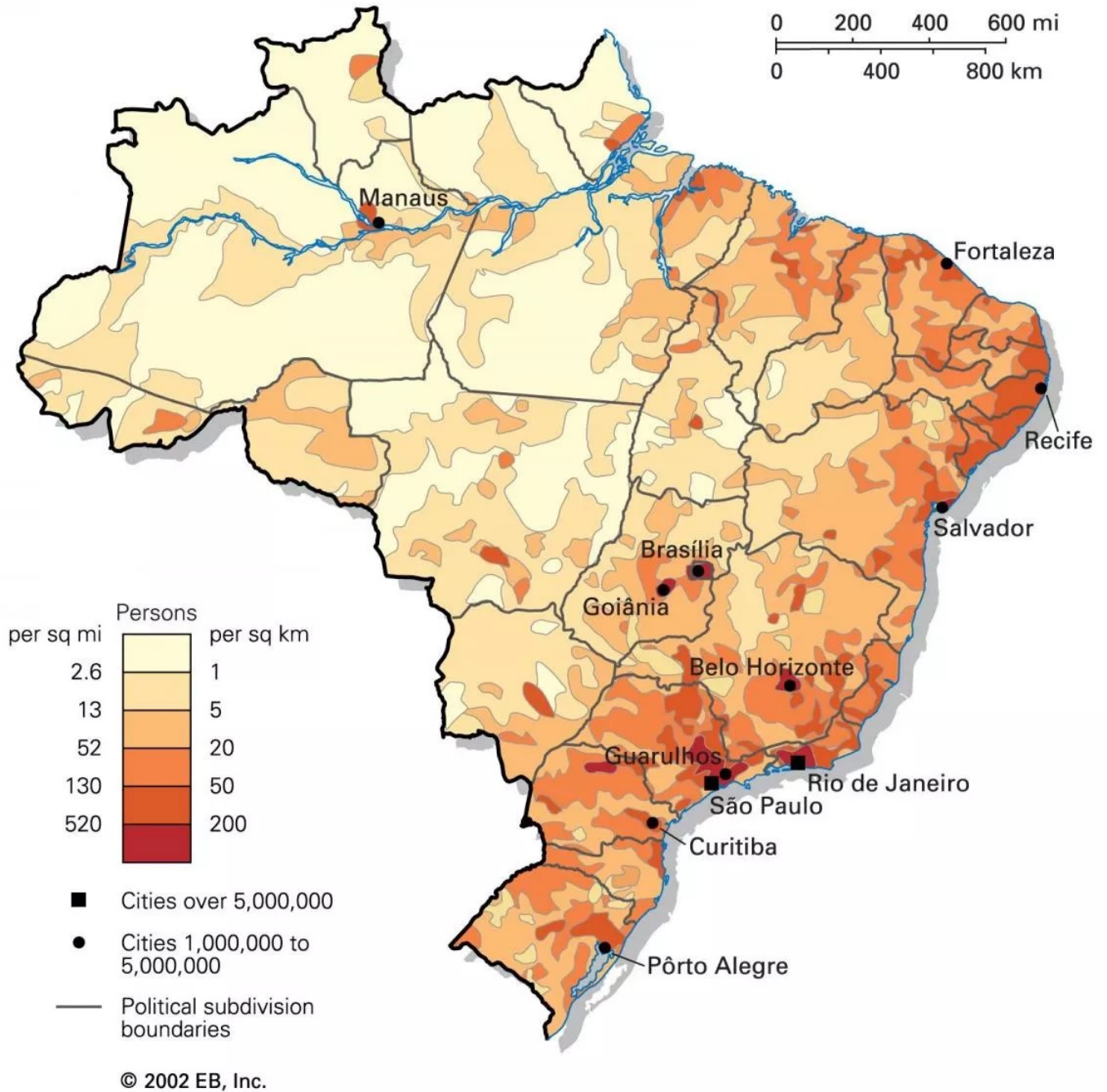


Image 2.

The image below shows the administrative map of Brazil.



> Actionable Insights and Recommendations

The following are the actionable insights that can be drawn from the analysis:

1. All the order purchases were made during the Morning, Afternoon and Evening time. And there were no orders made during the night, this is probably because a customer cannot place an order after the store timings are over, which according to the data is at 9 PM. One recommendation for this would be, to keep the online store open even after the physical store has closed for the day. That way, the customers that prefer to make purchases online and have the order delivered to them, can do so from the comfort of their homes.
2. Brazilians usually shop during the time of their festivities, this is quite evident from the data, which is from the month of January to August (both included). To boost sales during the remainder of the year, one recommendation would be to provide the customers with attractive offers.
3. A 40.65% increase in sales is observed from 2017 to 2018 for the time duration from January to August (both included).
4. The Northern part of Brazil which has the states of Amazonia, Roraima, Acre, Para, Amapa and Rondonia has the highest average freight value, highest average time of delivery and slowest average delivery time. This is because:
 1. This region is not densely populated.
 2. This region does not have a major port.
 3. And finally, it seems that it is also not well connected with the rest of the country.

This problem may be solved by looking at options across the border, into the bordering countries, whose ports are closer to these regions than the ones which are in Brazil. That way, the time to transport the orders may reduce, and this region may also show growth in sales.

> Reference

1. Festivals Celebrated in Brazil: <https://traveltriangle.com/blog/brazil-festivals/>
2. Brazilian States & Abbreviations: https://brazil-help.com/brazilian_states.htm
3. Population Density Map of Brazil: <https://brazilmap360.com/brazil-population-map>
4. Administrative Map of Brazil: <https://www.nationsonline.org/oneworld/map/brazil-administrative-map.htm>