Target – Brazil Market Case Study, Insights & Recommendations

Submitted by: Abhishek Singh Date: March 7,2023

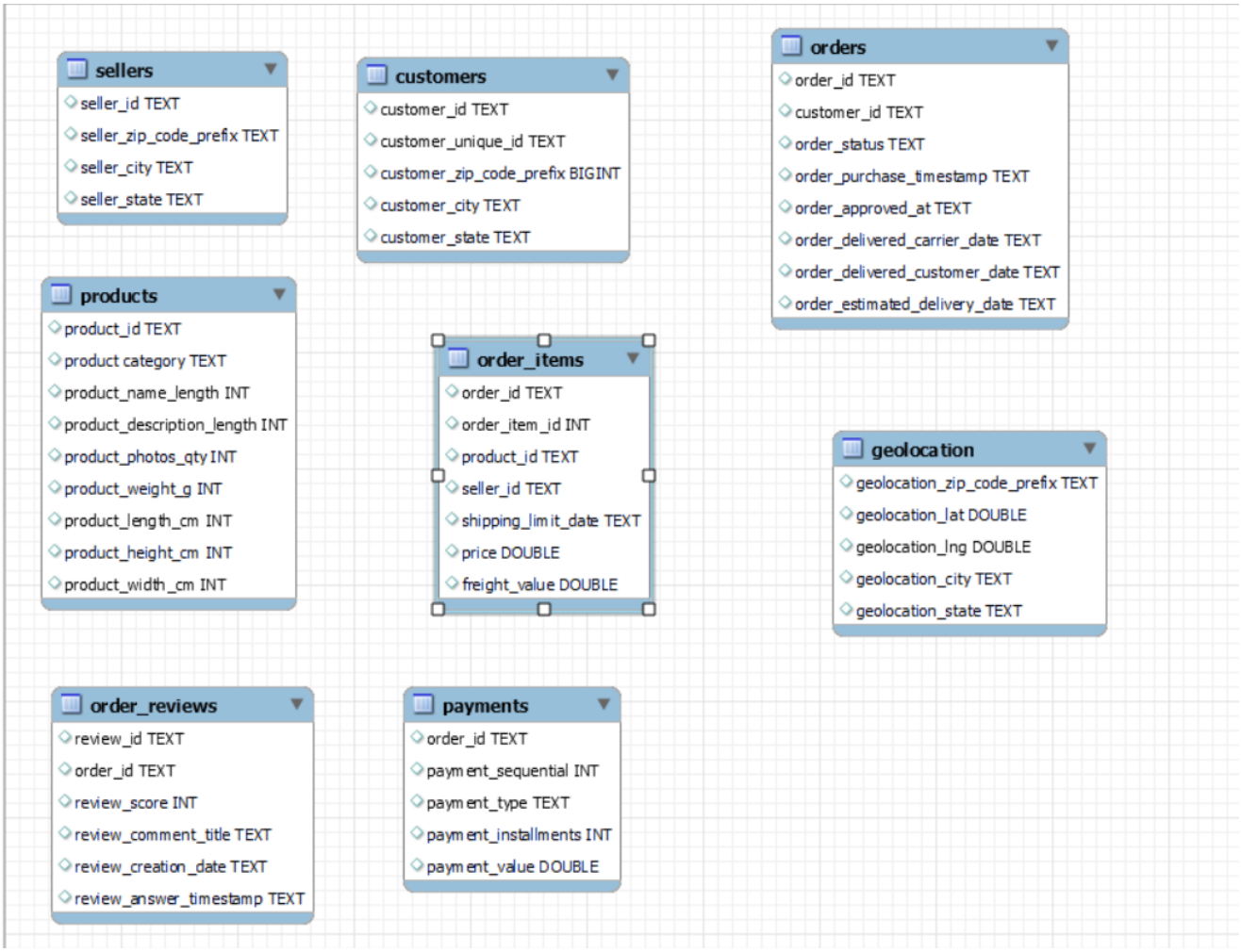
Objective: To gain insights about the shopping trends, sales & revenue generated, profit/loss analysis about the business operation of Target for the market and period provided and provide recommendations from the insights.

Pre-requisites: The data provided in CSV files were imported to a MySQL workbench for analysis and generating insights. CSV file names were used as the SQL table names and the steps to import large datasets efficiently can be referred here.

Initial Analysis & Observations:

1. Figuring out the metadata of the tables

   Approach: We generated an ER diagram in MySQL workbench to understand the metadata of all the tables. This ER diagram would also help us understand the kind of data in each table



Observations:

Table Name	Column Name	Column Description
customers	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 the location of the consumer.
	customer_city	Name of the City from where order is made.
	customer_state	State Code from where order is made(Ex- sao paulo-SP).
sellers	seller_id	Unique Id of the seller registered
	seller_zip_code_prefix	Zip Code of the location of the seller.
	seller_city	Name of the City of the seller.
	seller_state	State Code (Ex- sao paulo-SP)
order_items	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 Target.
	shipping_limit_date	The date before which shipping of the ordered product must be completed.
	price	Actual price of the products ordered .
	freight_value	Price rate at which a product is delivered from one point to another.
geolocation	geolocation_zip_code_prefix	first 5 digits of zip code
	geolocation_lat	latitude
	geolocation_lng	longitude
	geolocation_city	city name
	geolocation_state	state
payments	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.(Ex-Credit Card)
	payment_installments	number of installments in case of EMI purchase.
	payment_value	Total amount paid for the purchase order.
orders	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.
order_reviews	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 the 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.
	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 ordered.
	product_description_lenght	length of the description written for each product ordered on the site.
	product_photos_qty	Number of photos of each product ordered available on the shopping portal.
	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.
products	product_width_cm	width of the product ordered in centimeters.

2. Time Period for which the data is given

Approach:

- From the ER diagram generated, we observe that the 'order_purchase_timestamp' column from 'orders' table stores the timestamp of each order purchased. Using min() and max() on the column we can get the time-period for which we have the data.
- To understand how data is stored in the column we pull 10 records from the column.

```
4
5 •  select distinct order_purchase_timestamp from orders limit 10;
```

esult Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch r

order_purchase_timestamp
2017-10-02 10:56:33
2018-07-24 20:41:37
2018-08-08 08:38:49
2017-11-18 19:28:06
2018-02-13 21:18:39
2017-07-09 21:57:05
2017-04-11 12:22:08
2017-05-16 13:10:30
2017-01-23 18:29:09
2017-07-29 11:55:02

- By extracting the date component from timestamp and pulling the max and min timestamp values we can get the time-period of data

```
5 •  select date(max(order_purchase_timestamp)) as Max_date,
6       date(min(order_purchase_timestamp)) as Min_date from orders;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Max_date	Min_date
2018-10-17	2016-09-04

Observation: The data has been provided for orders places between September2016 and October 2018

3. <u>Cities and States of customers ordered during the given period</u>

Approach: The 'geolocation_city' and 'geolocation_state' columns in geolocation table will contain the required information

Observations: Maximum customers are from states SP, MG and RJ and the states RR and AP has smallest consumer base, which indicates that efforts need to be made in SP, MG and RJ to ensure customer retention and strategies need to be developed to penetrate deeper in the RR and AP markets.

```
5 •  select  geolocation_state as state, count(geolocation_state) as count_state
6    from geolocation group by geolocation_state
7    order by count_state limit 10;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

state	count_state
RR	646
AP	853
AC	1301
AM	2432
RO	3478
SE	3563
TO	3576
AL	4183
PI	4549
RN	5041

```
5 •  select  geolocation_state as state, count(geolocation_state) as count_state
6    from geolocation group by geolocation_state
7    order by count_state desc limit 10;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

state	count_state
SP	404268
MG	126336
RJ	121169
RS	61851
PR	57859
SC	38328
BA	36045
GO	20139
ES	16748
PE	16432

4. <u>Customer Retention Percent</u>**:**

SQL Script:

```
select
count(distinct customer_id) as cust_count,
count(distinct customer_unique_id) as unique_cust_count,
round((count(distinct customer_id)-count(distinct customer_unique_id))/count(distinct
customer_id)*100,2) as cust_retention
from customers;
```

cust_count	unique_cust_count	cust_retention
99441	96096	3.36

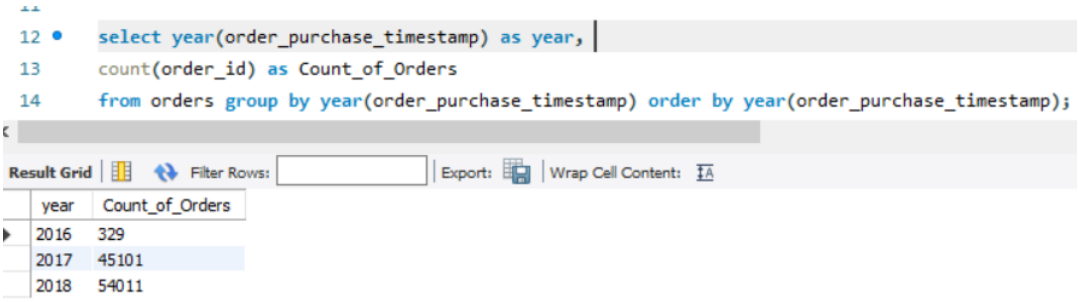
E-Commerce Trends in Brazil:

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?

   Approach: We can gain insights on this business problem by checking how the MOM and YOY growth in sales and revenue in the region

   Observations: We can observe that there has been a YOY growth in the number of orders places in Target across Brazil since 2016. From 2016 to 2017, there was a exponential growth owing to large untapped market in Brazil, however between 2017 to 2017 the decrease in growth was sharper than expected which could indicates an untapped population which can be leveraged for future growth, especially in states with least contribution to total orders placed

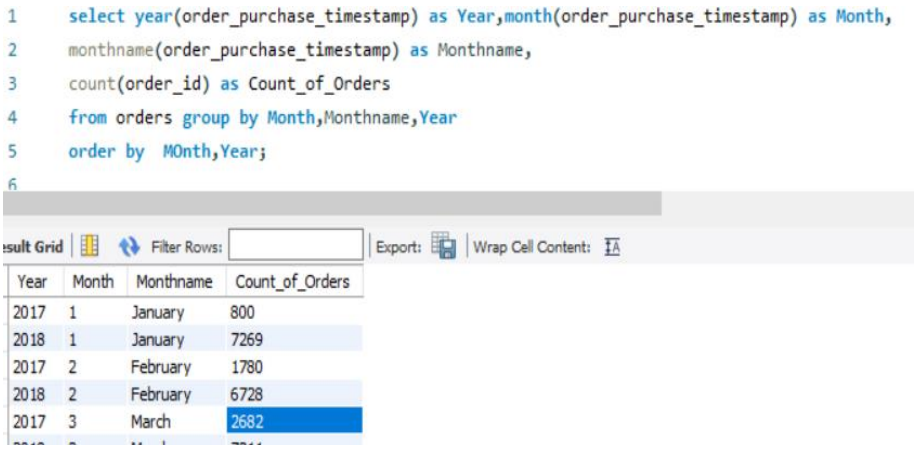
   A Month-on-Month analysis suggests that Christmas and New-year period (November to January) has the highest orders places each year, and incentives should be rolled out encouraging people to order more frequently during the months of Mar-Jun and Aug-Sep which have consistently seen a decline in orders placed each year.



```
12 • select year(order_purchase_timestamp) as year,
13   count(order_id) as Count_of_Orders
14   from orders group by year(order_purchase_timestamp) order by year(order_purchase_timestamp);
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

year	Count_of_Orders
2016	329
2017	45101
2018	54011

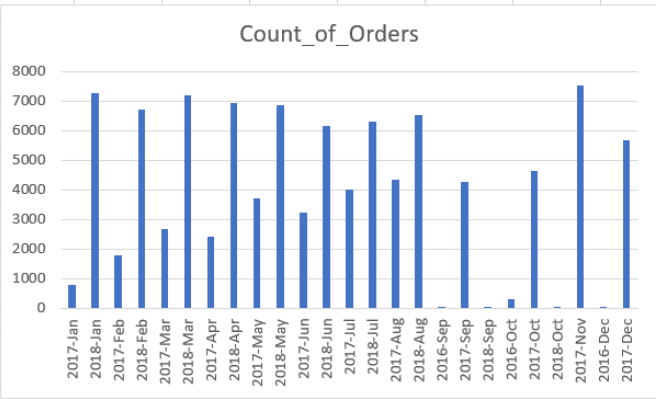


```
1  select year(order_purchase_timestamp) as Year,month(order_purchase_timestamp) as Month,
2  monthname(order_purchase_timestamp) as Monthname,
3  count(order_id) as Count_of_Orders
4  from orders group by Month,Monthname,Year
5  order by  MOnth,Year;
6
```

esult Grid | Filter Rows: | Export: | Wrap Cell Content:

Year	Month	Monthname	Count_of_Orders
2017	1	January	800
2018	1	January	7269
2017	2	February	1780
2018	2	February	6728
2017	3	March	2682

Period	Shopping trends based on number of orders placed
Feb - Mar	Increase in orders placed
Mar- Jun	Decrease in orders placed
Jun - Aug	Sharp increase in orders placed
Aug - Sep	Decrease in orders placed
Nov, Dec, Jan	Highest number of orders placed



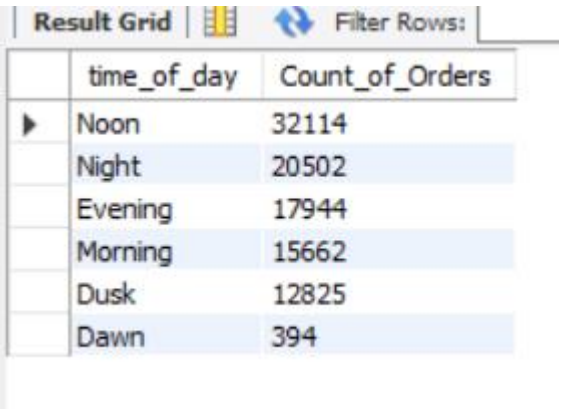
2. What time do Brazilian customers tend to buy (Dawn, Morning, Afternoon or Night)?

Approach: We can observe this trend by analyzing the orders placed at various hours of the day.

Observations: Shopping trends suggest that maximum number of orders are placed during Noon and Night, using which we can infer advertisements should be focused during this time to keep the revenue high during this day. Additionally, the least orders are placed during Dawn time, which can be accounted to fewer number of people using their PDAs during this time and consequently pulling off advertisements during this time might be a safe idea.

SQL Script:

```
select
case
when hour(order_purchase_timestamp) between 21 and 23
OR hour(order_purchase_timestamp) between 00 and 03 then 'Night'
when hour(order_purchase_timestamp) between 04 and 05 then 'Dawn'
when hour(order_purchase_timestamp) between 06 and 10 then 'Morning'
when hour(order_purchase_timestamp) between 11 and 15 then 'Noon'
when hour(order_purchase_timestamp) between 16 and 17 then 'Dusk'
when hour(order_purchase_timestamp) between 18 and 20 then 'Evening'
end as time_of_day,
count(order_id) as Count_of_Orders
from orders group by time_of_day
order by Count_of_Orders desc
```



time_of_day	Count_of_Orders
Noon	32114
Night	20502
Evening	17944
Morning	15662
Dusk	12825
Dawn	394

Evolution of E-commerce orders in the Brazil region:

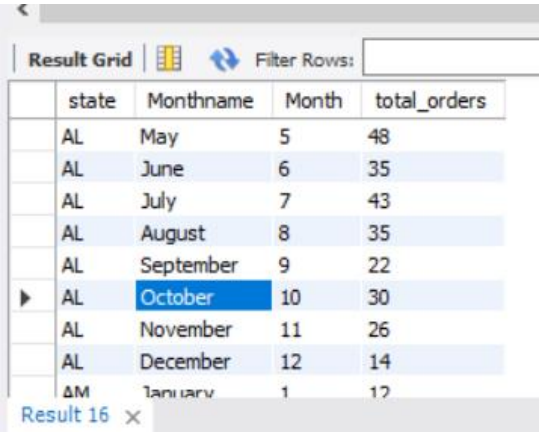
1. Get month on month orders by states

Approach: The insights can be determined by observing the total number of orders that have been placed across all the states for each month

Observations: We observe that the highest number of orders have been placed in August, in line with our observation above, and the state SP has highest contribution to the orders, and least being AC, indicating that the AC market has not been penetrated yet, and capital expenditure should be allocated to acquire a larger section of AC market

SQL Script:

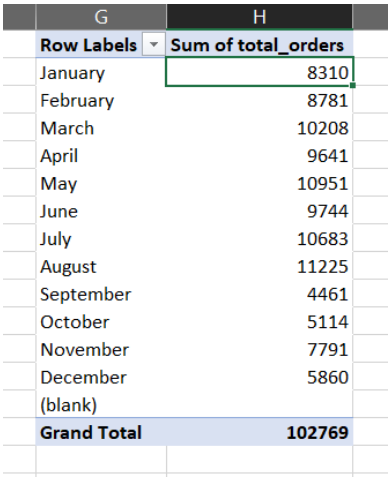
```
select
c.customer_state as state,
monthname(o.order_purchase_timestamp) as Monthname,
month(o.order_purchase_timestamp) as Month,
count(*) as total_orders
from orders o
join customers c on c.customer_id=o.customer_id
group by state, monthname, month
order by state, month;
```



Result Grid | Filter Rows:

state	Monthname	Month	total_orders
AL	May	5	48
AL	June	6	35
AL	July	7	43
AL	August	8	35
AL	September	9	22
AL	October	10	30
AL	November	11	26
AL	December	12	14
AM	January	1	12

Result 16



Row Labels	Sum of total_orders
January	8310
February	8781
March	10208
April	9641
May	10951
June	9744
July	10683
August	11225
September	4461
October	5114
November	7791
December	5860
(blank)	
Grand Total	**102769**

2. Distribution of customers across the states in Brazil:

Approach: We shall be plotting the number of customers across the states and consequently the number of orders placed per state

Observations: State SP has highest number of orders placed and the highest number of customers while state RR is yet to be tapped completely with lesser than 100 registered customers

SQL Script:

```
select
c.customer_state as state,
count(o.order_id) as total_orders
from orders o
join  customers c on c.customer_id=o.customer_id
group by state
order by total_orders desc;
```

```
select
customer_state as state,
count(distinct customer_unique_id) as cust_count
from customers
group by state
order by cust_count desc;
```

state	cust_count
SP	40302
RJ	12384
MG	11259
RS	5277
PR	4882
SC	3534
BA	3277
DF	2075
ES	1964
GO	1952
PE	1609
CE	1313
PA	949
MT	876

state	total_orders
SP	43208
RJ	13284
MG	12007
RS	5643
PR	5194
SC	3756
BA	3489
DF	2214
ES	2107
GO	2097
PE	1698
CE	1378
PA	1001
MT	945

Impact on Economy:

1. Get % increase in cost of orders from 2017 to 2018:

   Approach: We shall compute the total payments received across the two years

   Observations: The total cost of orders in Brazil market from 2017 to 2018 has increased by 136%, indicating an overall good brand establishment and gradual increase in loyalty towards business by the customers, resulting in overall good peer-to-peer advertisement and growth

Result Grid | Filter Rows:

cost_of_orders	year
3669022.12 Brazil Real	2017
8694733.84 Brazil Real	2018

SQL Script:

```
select
concat(round(sum(payment_value),2)," Brazil Real") as cost_of_orders,
year(o.order_purchase_timestamp) as year
from orders o
join payments p on o.order_id=p.order_id
where month(o.order_purchase_timestamp) between 1 and 8
and year(o.order_purchase_timestamp) in (2017,2018)
group by year
order by year;
```

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

   Approach: Compute the prices

   Observations: For the 10 states in the output, we observe that the average price of orders is high despite the average freight price being high as well. Establishing better logistic network in these 10 states to bring down the overall freight cost, could have positive impact on the consumer purchasing parity and result in higher sales.

total_price	avg_price	total_freight	avg_freight	cust_state
116225.99	189.6	26303.2	42.91	PB
82634.51	180.82	16439.85	35.97	AL
16224.74	170.79	3769.85	39.68	AC
47014.53	166.72	11578.74	41.06	RO
13813.2	164.44	2916.34	34.72	AP
182482.1	164.4	39621.18	35.69	PA
90981.33	161.89	22002.45	39.15	PI
52036.23	161.6	11945.74	37.1	TO
8629.13	156.89	2448.64	44.52	RR
84710.6	156	19403.95	35.73	RN

SQL Script:

```
select
round(sum(oi.price),2) as total_price,
round(avg(oi.price),2) as avg_price,
round(sum(freight_value),2) as total_freight,
round(avg(freight_value),2) as avg_freight,
c.customer_state as cust_state
from orders o
join order_items oi on o.order_id=oi.order_id
join customers c on c.customer_id=o.customer_id
group by c.customer_state
order by avg_price desc, avg_freight desc;
```

Analysis on sales, freight, and delivery time:

1. Calculate days between purchasing, delivering and estimated delivery:

   Observations: The average delivery time for state RR is the highest, while the total orders placed is the least among all the states. The very high delivery times could be a reason for customers to turn away from placing orders, which can be remediated by investing in logistics in the state to attract more customers and ensure future growth.

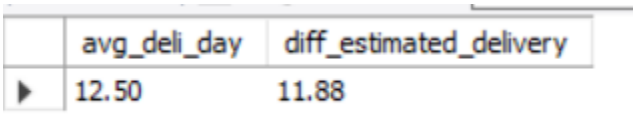
SQL Script:

```
select
round(avg(datediff(o.order_delivered_customer_date,date(o.order_purchase_timestamp))),2) as
avg_deli_day,
round(avg(datediff(o.order_estimated_delivery_date,date(o.order_purchase_timestamp))),2) as
avg_est_deli_day,
c.customer_state as cust_state,
count(o.order_id) as total_orders
from orders o
join customers c on o.customer_id=c.customer_id
group by cust_state
order by avg_deli_day desc, avg_est_deli_day desc, total_orders desc;
```

avg_deli_day	avg_est_deli_day	cust_state	total_orders
29.16	47.40	RR	48
27.42	46.79	AP	70
26.42	45.78	AM	154
24.79	33.18	AL	425
23.67	37.72	PA	1001
21.55	31.05	MA	766
21.48	31.35	SE	360
21.22	31.96	CE	1378
21.00	41.56	AC	84
20.37	33.69	PB	546
19.37	30.80	PI	511
19.31	39.39	RO	257

2. Find time_to_delivery & diff_estimated_delivery:

Observations: The orders are delivered on earlier than the estimated delivery date on average.



avg_deli_day	diff_estimated_delivery
12.50	11.88

SQL Script:

```
select
round(avg(datediff(o.order_delivered_customer_date,date(o.order_purchase_timestamp))),2) as avg_deli_day,
round(avg(datediff(o.order_estimated_delivery_date,o.order_delivered_customer_date)),2) as diff_estimated_delivery
from orders o;
```

3. Group data by state, take mean of freight_value, time_to_delivery, diff_estimated_delivery:

SQL Script:

```
select
c.customer_state as cust_state
round(avg(datediff(o.order_delivered_customer_date,date(o.order_purchase_timestamp))),2) as avg_deli_day,
round(avg(datediff(o.order_estimated_delivery_date,o.order_delivered_customer_date)),2) as diff_estimated_delivery,
round(avg(oi.freight_value),2) as avg_freight_value,
from orders o
join order_items oi on o.order_id=oi.order_id
join customers c on c.customer_id=o.customer_id
group by c.customer_state
```

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

Observation: The highest and lowest average freight value belong to the state with lowest and highest contribution to total orders respectively. Freight charges affect the customer purchasing decision negatively.

SQL Script:

```
select
c.customer_state as cust_state
round(avg(datediff(o.order_delivered_customer_date,date(o.order_purchase_timestamp))),2) as avg_deli_day,
round(avg(datediff(o.order_estimated_delivery_date,o.order_delivered_customer_date)),2) as diff_estimated_delivery,
round(avg(oi.freight_value),2) as avg_freight_value,
from orders o
join order_items oi on o.order_id=oi.order_id
join customers c on c.customer_id=o.customer_id
group by c.customer_state
order by avg_freight_value desc
limit 5;
```

	cust_state	avg_deli_day	diff_estimated_delivery	avg_freight_value
▶	RR	28.10	18.57	44.52
	PB	20.53	13.16	42.91
	RO	19.67	20.00	41.06
	AC	20.69	20.79	39.68
	PI	19.31	11.57	39.15

SQL Script:

```
select
c.customer_state as cust_state
round(avg(datediff(o.order_delivered_customer_date,date(o.order_purchase_timestamp))),2) as avg_deli_day,
round(avg(datediff(o.order_estimated_delivery_date,o.order_delivered_customer_date)),2) as diff_estimated_delivery,
round(avg(oi.freight_value),2) as avg_freight_value,
from orders o
join order_items oi on o.order_id=oi.order_id
join customers c on c.customer_id=o.customer_id
group by c.customer_state
order by avg_freight_value asc
limit 5;
```

	cust_state	avg_deli_day	diff_estimated_delivery	avg_freight_value
▶	SP	8.67	11.20	15.15
	PR	11.90	13.48	20.58
	MG	11.91	13.35	20.62
	RJ	15.06	12.02	20.93
	DF	12.86	12.22	21.03

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

Observation: Shopping trends suggest that the delivery time impacts the purchasing decision of customer inversely. States with low delivery times witnessed the highest number of orders placed and vice-versa

SQL Script:

```
select
c.customer_state as cust_state,
round(avg(datediff(o.order_delivered_customer_date,date(o.order_purchase_timestamp))),2) as avg_deli_day,
round(avg(datediff(o.order_estimated_delivery_date,o.order_delivered_customer_date)),2) as diff_estimated_delivery,
round(avg(oi.freight_value),2) as avg_freight_value
from orders o
join order_items oi on o.order_id=oi.order_id
join customers c on c.customer_id=o.customer_id
group by c.customer_state
order by avg_deli_day desc
limit 5;
```

cust_state	avg_deli_day	diff_estimated_delivery	avg_freight_value
AP	28.40	18.29	34.72
RR	28.10	18.57	44.52
AM	26.39	19.88	33.02
AL	24.72	8.40	35.97
PA	23.62	14.28	35.69

SQL Script:

```
select
c.customer_state as cust_state,
round(avg(datediff(o.order_delivered_customer_date,date(o.order_purchase_timestamp))),2) as avg_deli_day,
round(avg(datediff(o.order_estimated_delivery_date,o.order_delivered_customer_date)),2) as diff_estimated_delivery,
round(avg(oi.freight_value),2) as avg_freight_value
from orders o
join order_items oi on o.order_id=oi.order_id
join customers c on c.customer_id=o.customer_id
group by c.customer_state
order by avg_deli_day asc
limit 5;
```

cust_state	avg_deli_day	diff_estimated_delivery	avg_freight_value
SP	8.67	11.20	15.15
PR	11.90	13.48	20.58
MG	11.91	13.35	20.62
DF	12.86	12.22	21.03
SC	14.98	11.55	21.4

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

Observation: As per the data, for the state with highest difference between estimated delivery date and actual delivery date, the customer purchase decision was primarily impacted by the delivery dates, which imply the delivery date estimation should be improved in order to encourage more customer turnup in the affected state.

SQL Script:

```
select
c.customer_state as cust_state,
round(avg(datediff(o.order_delivered_customer_date,date(o.order_purchase_timestamp))),2) as avg_deli_day,
round(avg(datediff(o.order_estimated_delivery_date,o.order_delivered_customer_date)),2) as diff_estimated_delivery,
round(avg(oi.freight_value),2) as avg_freight_value
from orders o
join order_items oi on o.order_id=oi.order_id
```

```
join customers c on c.customer_id=o.customer_id
group by c.customer_state
order by diff_estimated_delivery desc
limit 5;
```

cust_state	avg_deli_day	diff_estimated_delivery	avg_freight_value
AP	28.40	18.29	34.72
RR	28.10	18.57	44.52
AM	26.39	19.88	33.02
AL	24.72	8.40	35.97
PA	23.62	14.28	35.69

SQL Script:

```
select
c.customer_state as cust_state,
round(avg(datediff(o.order_delivered_customer_date,date(o.order_purchase_timestamp))),2) as
avg_deli_day,
round(avg(datediff(o.order_estimated_delivery_date,o.order_delivered_customer_date)),2) as
diff_estimated_delivery,
round(avg(oi.freight_value),2) as avg_freight_value
from orders o
join order_items oi on o.order_id=oi.order_id
join customers c on c.customer_id=o.customer_id
group by c.customer_state
order by diff_estimated_delivery asc
limit 5;
```

cust_state	avg_deli_day	diff_estimated_delivery	avg_freight_value
AL	24.72	8.40	35.97
MA	21.63	9.80	38.25
SE	21.46	9.94	36.92
ES	15.63	10.62	22.11
BA	19.16	11.04	26.31

Payment type analysis:

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

Observation: Majority of customers prefer to pay via UPI, followed by Credit cards. Establishing partnerships with major banking institutions to provide credit card offers upon performing

total_orders	Month	Monthname	Pay_type
6103	1	January	credit_card
118	1	January	debit_card
477	1	January	voucher
1715	1	January	UPI
82	2	February	debit_card
1723	2	February	UPI
6609	2	February	credit_card
424	2	February	voucher
7707	3	March	credit_card
591	3	March	voucher
1942	3	March	UPI
109	3	March	debit_card

Row Labels	Sum of # total_orders
credit_card	76795
debit_card	1529
not_defined	3
UPI	19784
voucher	5775
(blank)	
Grand Total	**103886**

SQL Script:

```
select
count(p.order_id) as total_orders,
month(o.order_purchase_timestamp) as Month,
monthname(o.order_purchase_timestamp) as Monthname,
p.payment_type as Pay_type
from orders o
join payments p on o.order_id=p.order_id
group by Pay_type, Month, Monthname
order by Month;
```

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

Observations: Majority of customers prefer purchasing items in a single installment. Reluctance of choosing multiple installments could result in customers shying away from major/bulk purchases. Enabling zero cost EMI options could encourage impact this positively.

SQL Script:

```
select
count(p.order_id) as total_orders,
p.payment_installments as Installments
from payments p
group by Installments
order by total_orders desc;
```

total_orders	Installments
52546	1
12413	2
10461	3
7098	4
5328	10
5239	5
4268	8

Actionable Insights:

- Though boasting a large customer base, by comparing the total number of customer id with the unique customer ids we can conclude that the customer retention stands at 3.36% which suggests that customers do not return for repeat purchases. This can be further concluded by the exponential rise in customer base and subsequently orders placed from 2016 to 2017 and then the drastic decline in total orders placed from 2017 to 2018.
- Markets that are currently not yet tapped completely are due to a combination of high freight charges, high estimated and actual delivery time.
- Credit card utilization for purchases is low, despite being the strongest financial instrument.
- Majority of customers are saturated in the states SP, RJ and MG and consequently these states have highest contribution to the total orders. Markets in RR, AM and AC are yet to be utilized to the full potential and capital expenditure for market penetration should be considered.
- Bulk and expensive purchases should be expected and prepared for in the months of July and August, since these two months witness largest shopping spree.

Recommendations:

- Rollout of loyalty/membership program to improve customer retention and encourage multiple orders by customers.
- Establish partnership with local logistics providers to improve freight charges and delivery times in states with worst delivery times.
- Partnership with financial institutions to roll out credit card offers and no cost EMIs plans for customers to improve credit card utilization and encourage larger purchases.
- Capital expenditure to roll out discount offers and free delivery on purchases to state with a non-established/minute customer base.
- Enhance advertisement during the afternoon and night-time to promote higher profit margin products, as these time periods of the day are the segment leaders and bring in largest volume of orders.