
Business Case: Target SQL

Introduction

Target is one of the most recognizable brands in the world and a top retailer in the United States. By providing unmatched value, creativity, innovation, and an extraordinary customer experience that no other retailer can match, Target establishes itself as a preferred shopping destination. This business case contains data on 100k orders placed at Target in Brazil between 2016 and 2018. Its capabilities enable examining an order from a variety of angles, including customer location, product qualities, order status, pricing, payment, and freight performance.

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

```
Select column_name, data_type
from `target-case-study-1.target_retail.INFORMATION_SCHEMA.COLUMNS`
where table_name = 'customers';
```

Query results			
JOB INFORMATION		RESULTS	JSON
Row	column_name	data_type	
1	customer_id	STRING	
2	customer_unique_id	STRING	
3	customer_zip_code_prefix	INT64	
4	customer_city	STRING	
5	customer_state	STRING	

Insight:

- It appears that the Customers table has 4 out of 5 string features(customer_id, customer_unique_id, customer_city, customer_state) and only 1 numerical feature (customer_zip_code_prefix).

2. Time period for which the data is given

```
Select
(Date_Diff(max_pur,min_pur, year)+1) as Time_Period_Year
from
(
Select Min(date(order_purchase_timestamp)) as min_pur,
Max(date(order_purchase_timestamp)) as max_pur
from `target_retail.orders`);
```

Query results		
JOB INFORMATION		RESULTS
Row	Time_Period_Year	
1	3	

Insight:

- It seems that the data given in the Target dataset is for 3 years.

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

```
select distinct customer_city,  
customer_state  
from `target_retail.customers`;
```

Query results			
JOB INFORMATION		RESULTS	JSON
Row	customer_city	customer_state	EXECUTION DET
1	acu	RN	
2	ico	CE	
3	ipe	RS	
4	ipu	CE	
5	ita	SC	
6	itu	SP	
7	jau	SP	
8	luz	MG	
9	poa	SP	
10	uba	MG	

Insight:

- In the given dataset, the orders have been received from 50 distinct cities in different states of Brazil.
- Further analysis reveals that Targets' client base is drawn from all 27 Brazilian states.

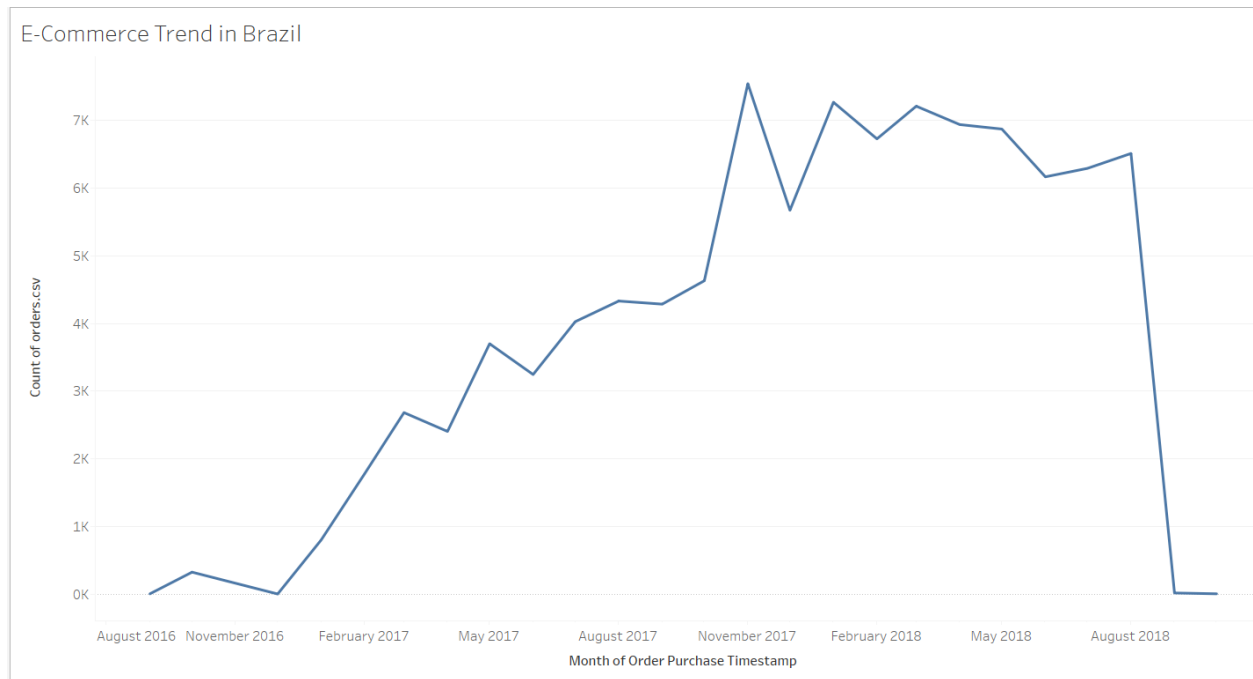
Question 2:

In-depth Exploration:

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

```
Select Count(order_id) as No_of_orders,  
EXTRACT (Year from order_purchase_timestamp) as Year,  
EXTRACT (Month from order_purchase_timestamp) as Month  
from `target_retail.orders`  
Group by Year, Month  
Order by Year, Month;
```

JOB INFORMATION		RESULTS		JSON
Row	No_of_orders	Year	Month	
1	4	2016	9	
2	324	2016	10	
3	1	2016	12	
4	800	2017	1	
5	1780	2017	2	
6	2682	2017	3	
7	2404	2017	4	
8	3700	2017	5	
9	3245	2017	6	
10	4026	2017	7	



Insights:

- With the help of the above plot and the data, we can observe that there is an increase in the number of customer orders from August 2016 to November 2017 which is when the number of orders were the highest.
- Between November 2017 to August 2018, the number of orders placed, appears to have a steady situation until September 2018 when it drops drastically.

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

```
Select Count(order_id) as orders_purchased,time_period
from
(
Select order_id,
Case When extract(hour from order_purchase_timestamp)Between 0 And 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 time_period
from `target_retail.orders`) o
Group by time_period
order by orders_purchased desc;
```

JOB INFORMATION		RESULTS
Row	orders_purchased	time_period
1	38135	Afternoon
2	28331	Night
3	27733	Morning
4	5242	Dawn

Insights:

- According to the data presented above, Brazilian buyers tend to buy at all times (Dawn, Morning, Afternoon, and Night).
- However, the most orders are placed in the afternoon, while the least number of orders are placed during the dawn.

Question 3:

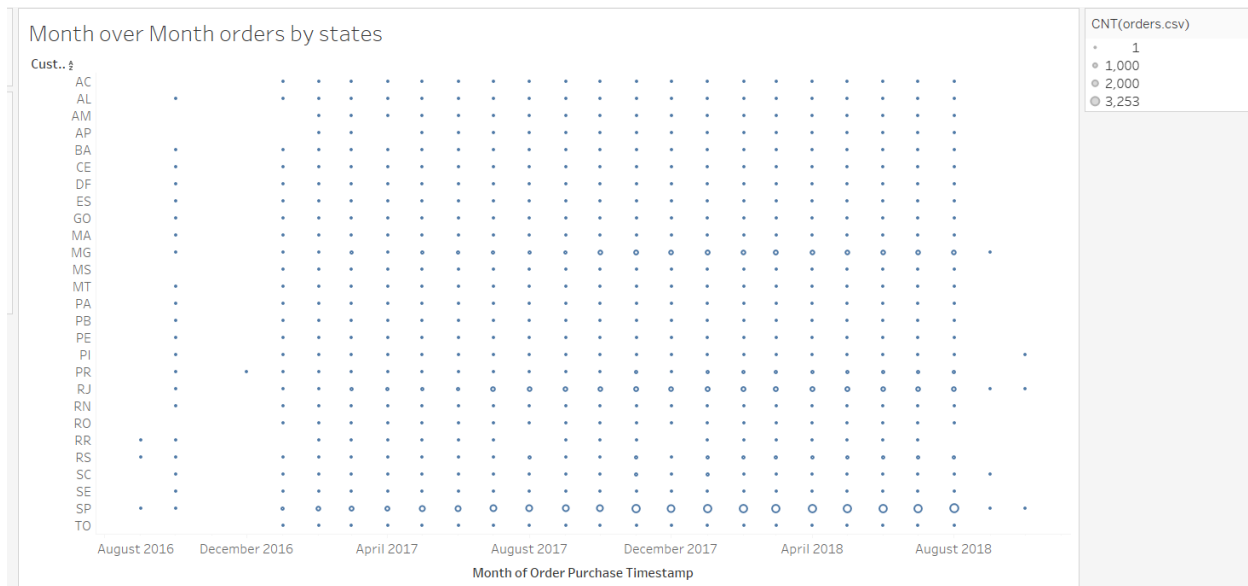
Evolution of E-commerce orders in Brazil region:

1. Get month on month orders by states

```
Select Count(order_id) as No_of_orders, customer_state,  
Extract(year from order_purchase_timestamp) as Year,  
Extract(month from order_purchase_timestamp) as Month  
from `target_retail.orders` o inner join `target_retail.customers` c  
ON c.customer_id = o.customer_id  
Group by customer_state,Year,Month  
Order by Year, Month;
```

Query results

JOB INFORMATION		RESULTS	JSON	EXECUTION DETAILS		EXEC
Row	No_of_orders	customer_state		Year	Month	
1	1	RR		2016	9	
2	1	RS		2016	9	
3	2	SP		2016	9	
4	113	SP		2016	10	
5	24	RS		2016	10	
6	56	RJ		2016	10	
7	3	MT		2016	10	
8	9	GO		2016	10	
9	40	MG		2016	10	
10	8	CE		2016	10	



Insight:

- The month-on-month orders by state data demonstrates the total number of orders placed in different Brazilian states month by month in these 3-year timespan.
- It can be observed that the three states with the maximum count of orders appear to be 'MG', 'RJ', 'SP', where SP has the highest count of orders among all other states and it is increasing month over month from December 2016 until August 2018.

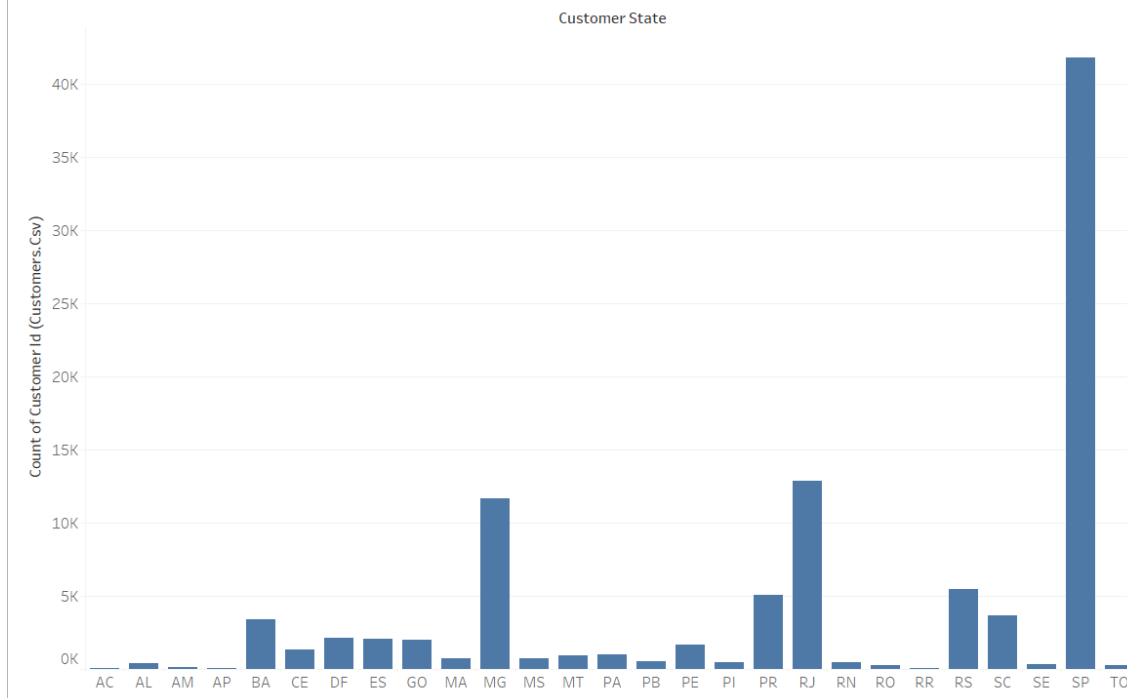
2. Distribution of customers across the states in Brazil

```
Select customer_state, count(customer_id)as No_of_customers
from `target_retail.customers`
Group by customer_state
Order by No_of_customers desc;
```

Query results

JOB INFORMATION		RESULTS	JSON
Row	customer_state	No_of_customer	
1	SP	41746	
2	RJ	12852	
3	MG	11635	
4	RS	5466	
5	PR	5045	
6	SC	3637	
7	BA	3380	
8	DF	2140	
9	ES	2033	
10	GO	2020	

Customers across Brazilian States



Insights:

- From the data and plot mentioned above, it can be seen that the state 'SP' has the largest customer base of Target in Brazil followed by 'RJ' and 'MG'.

Question 4:

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

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

```
With orders_2017 AS
(
  Select Extract(year from o.order_purchase_timestamp) as Prev_Year,
  Extract(month from o.order_purchase_timestamp) as Prev_Month,
  Sum(p.payment_value) as Total_payment_2017
  from
  `target_retail.orders` o inner join `target_retail.payments` p
  ON p.order_id = o.order_id
  Where Extract(year from o.order_purchase_timestamp) = 2017
  Group by Prev_Year, Prev_Month
  Order by Prev_Year, Prev_Month
),
orders_2018 AS
(
  Select Extract(year from o.order_purchase_timestamp) as Curr_Year,
  Extract(month from o.order_purchase_timestamp) as Curr_Month,
  Sum(p.payment_value) as Total_payment_2018
  from
  `target_retail.orders` o inner join `target_retail.payments` p
  ON p.order_id = o.order_id
  Where Extract(year from o.order_purchase_timestamp) = 2018
  Group by Curr_Year, Curr_Month
  Order by Curr_Year, Curr_Month
)
```

```

Select Prev_Year,Prev_Month, Curr_Year, Curr_Month,
(((Total_payment_2018 - Total_payment_2017)/Total_payment_2017)*100) as
Perc_inc
from orders_2017 a join orders_2018 b ON
a.Prev_Month = b.Curr_Month
where Prev_Month IN (1,2,3,4,5,6,7,8)
order by Prev_Month;

```

Query results						
JOB INFORMATION		RESULTS	JSON	EXECUTION DETAILS		EXECUTION GRAPH PRE
Row	Prev_Year	Prev_Month	Curr_Year	Curr_Month	Perc_inc	
1	2017	1	2018	1	705.1266954171...	
2	2017	2	2018	2	239.9918145445...	
3	2017	3	2018	3	157.7786066709...	
4	2017	4	2018	4	177.8407701149...	
5	2017	5	2018	5	94.62734375677...	
6	2017	6	2018	6	100.2596912456...	
7	2017	7	2018	7	80.04245463390...	
8	2017	8	2018	8	51.60600520477...	

Insights:

- It can be noted that the cost of orders has been increased for each month from 2017 to 2018 by some percentage.
- The highest percentage increase in the cost of orders from 2017 to 2018 has been observed for the month of January, whereas the lowest has been noted for August.

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

```
Select
Avg(o.price) as Mean_price,
Sum(o.price) as Total_price,
Avg(o.freight_value) as Mean_freight_value,
Sum(o.freight_value) as Total_frieght_value,
c.customer_state
from `target_retail.order_items` o inner join
`target_retail.orders` a on a.order_id = o.order_id
inner join `target_retail.customers` c on
c.customer_id = a.customer_id
Group by c.customer_state;
```

Query results

JOB INFORMATION		RESULTS	JSON	EXECUTION DETAILS		EXECUTION GRAPH	PREVIEW
Row		Mean_price ▼	Total_price ▼	Mean_freight_value	Total_frieght_value	customer_state ▼	
1		109.6536291597...	5202955.050002...	15.14727539041...	718723.0699999...	SP	
2		125.1178180945...	1824092.669999...	20.96092393168...	305589.3100000...	RJ	
3		119.0041393728...	683083.7600000...	20.53165156794...	117851.6800000...	PR	
4		124.6535775862...	520553.3400000...	21.47036877394...	89660.26000000...	SC	
5		125.7705486284...	302603.9399999...	21.04135494596...	50625.49999999...	DF	
6		120.7485741488...	1585308.029999...	20.63016680630...	270853.4600000...	MG	
7		165.6924166666...	178947.8099999...	35.83268518518...	38699.30000000...	PA	
8		134.6012082126...	511349.9900000...	26.36395893656...	100156.6799999...	BA	
9		126.2717316759...	294591.9499999...	22.76681525932...	53114.97999999...	GO	
10		120.3374530874...	750304.0200000...	21.73580433039...	135522.7400000...	RS	

Question 5:

Analysis on sales, freight, and delivery time

1. Calculate days between purchasing, delivering and estimated delivery

```
Select
order_id,
Date_diff(order_delivered_carrier_date, order_purchase_timestamp, Day) as
Days_to_Delivery,
Date_diff(order_estimated_delivery_date, order_purchase_timestamp, Day) as
Days_to_Estimated_Delivery
from
`target_retail.orders`;
```

Query results

JOB INFORMATION		RESULTS	JSON	EXECUTION DETAILS	EXECUTION TIME
Row	order_id	Days_to_Delivery	Days_to_Estimated_Delivery		
1	f88aac7ebccb37f19725a0753...	9	50		
2	790cd37689193dca0d00d2feb...	2	6		
3	49db7943d60b6805c3a41f547...	6	44		
4	063b573b88fc80e516aba87df...	22	54		
5	a68ce1686d536ca72bd2dad4...	33	56		
6	45973912e490866800c0aea8f...	18	54		
7	cda873529ca7ab71f677d5ec1...	39	56		
8	ead20687129da8f5d89d831bb...	1	41		
9	6f028ccb7d612af251aa442a1f...	1	3		
10	8733c8d440c173e524d2fab80...	0	3		

Insights:

- It appears that most of the orders have been delivered by the carrier much earlier as compared to the estimated delivery date.

2. Find time_to_delivery & diff_estimated_delivery. Formula for the same given below:

a. $\text{Time_to_delivery} = \text{order_delivered_customer_date} - \text{order_purchase_timestamp}$

```
select order_id, order_purchase_timestamp,  
order_delivered_customer_date,  
Date_diff(order_delivered_customer_date,order_purchase_timestamp, Day) as  
time_to_delivery  
from `target_retail.orders`;
```

Query results

JOB INFORMATION		RESULTS	JSON	EXECUTION DETAILS	EXECUTION GRAPH	PREVIEW
Row	order_id	order_purchase_timestamp	order_delivered_customer_date	time_to_delivery		
1	1950d777989f6a877539f5379...	2018-02-19 19:48:52 UTC	2018-03-21 22:03:51 UTC	30		
2	2c45c33d2f9cb8ff8b1c86cc28...	2016-10-09 15:39:56 UTC	2016-11-09 14:53:50 UTC	30		
3	65d1e226dfaeb8cdc42f66542...	2016-10-03 21:01:41 UTC	2016-11-08 10:58:34 UTC	35		
4	635c894d068ac37e6e03dc54e...	2017-04-15 15:37:38 UTC	2017-05-16 14:49:55 UTC	30		
5	3b97562c3aee8bdedcb5c2e45...	2017-04-14 22:21:54 UTC	2017-05-17 10:52:15 UTC	32		
6	68f47f50f04c4cb6774570cfde...	2017-04-16 14:56:13 UTC	2017-05-16 09:07:47 UTC	29		
7	276e9ec344d3bf029ff83a161c...	2017-04-08 21:20:24 UTC	2017-05-22 14:11:31 UTC	43		
8	54e1a3c2b97fb0809da548a59...	2017-04-11 19:49:45 UTC	2017-05-22 16:18:42 UTC	40		
9	fd04fa4105ee8045f6a0139ca5...	2017-04-12 12:17:08 UTC	2017-05-19 13:44:52 UTC	37		
10	302bb8109d097a9fc6e9cefc5...	2017-04-19 22:52:59 UTC	2017-05-23 14:19:48 UTC	33		

Insights:

- Time to delivery is at least 30 days or more for approximately all the orders that have been placed.
- The observation can be drawn with the help of this data shown above, that most customers have to wait for a month or over to receive the order that they placed.

b. $\text{diff_estimated_delivery} = \text{order_estimated_delivery_date} - \text{order_delivered_customer_date}$

```
select order_id, order_estimated_delivery_date,
order_delivered_customer_date,
Date_diff(order_estimated_delivery_date,order_delivered_customer_date, Day) as
diff_estimated_delivery
from `target_retail.orders`
order by diff_estimated_delivery desc;
```

Query results

JOB INFORMATION		RESULTS	JSON	EXECUTION DETAILS	EXECUTION GRAPH	PREVIEW
Row	order_id	order_estimated_delivery_date	order_delivered_customer_date	diff_estimated_delivery		
1	770d331c84e5b214bd9dc70a1...	2016-11-29 00:00:00 UTC	2016-10-14 15:07:11 UTC	45		
2	1950d777989f6a877539f5379...	2018-03-09 00:00:00 UTC	2018-03-21 22:03:51 UTC	-12		
3	2c45c33d2f9cb8ff8b1c86cc28...	2016-12-08 00:00:00 UTC	2016-11-09 14:53:50 UTC	28		
4	dabf2b0e35b423f94618bf965f...	2016-11-30 00:00:00 UTC	2016-10-16 14:36:59 UTC	44		
5	8beb59392e21af5eb9547ae1a...	2016-11-30 00:00:00 UTC	2016-10-19 18:47:43 UTC	41		
6	b60b53ad0bb7dacacf2989fe2...	2017-05-18 00:00:00 UTC	2017-05-23 13:12:27 UTC	-5		
7	276e9ec344d3bf029ff83a161c...	2017-05-18 00:00:00 UTC	2017-05-22 14:11:31 UTC	-4		
8	1a0b31f08d0d7e87935b819ed...	2017-05-18 00:00:00 UTC	2017-04-18 08:18:11 UTC	29		
9	cec8f5f7a13e5ab934a486ec9e...	2017-05-18 00:00:00 UTC	2017-04-07 13:14:56 UTC	40		
10	2d846c03073b1a424c1be1a77...	2017-05-18 00:00:00 UTC	2017-05-25 10:49:48 UTC	-7		

Insight:

- It can be seen that there is a difference between the order's actual delivery date compared to its estimated delivery date for almost every order.

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

```
select count(o.order_id) as count_of_orders,c.customer_state,
Avg(i.freight_value) as Mean_freight_value,
Avg(Date_diff(order_delivered_customer_date,order_purchase_timestamp, Day)) as
time_to_delivery,
Avg(Date_diff(order_estimated_delivery_date,order_delivered_customer_date, Day)) as
diff_estimated_delivery
from `target_retail.order_items` i inner join `target_retail.orders` o ON
o.order_id = i.order_id inner join `target_retail.customers` c ON
c.customer_id = o.customer_id
Group by o.order_id, c.customer_state;
```

Query results

JOB INFORMATION		RESULTS	JSON	EXECUTION DETAILS		EXECUTION GRAPH	PREVIEW
Row	count_of_orders	customer_state		Mean_freight_value	time_to_delivery	diff_estimated_delivery	
1	1	RJ		14.11	45.0	7.0	
2	2	SP		22.93	29.0	-8.0	
3	1	RJ		16.11	83.0	-56.0	
4	1	RJ		14.1	32.0	-7.0	
5	1	SP		12.79	30.0	-15.0	
6	1	RN		24.84	35.0	2.0	
7	1	RJ		18.23	43.0	-7.0	
8	1	RJ		15.23	42.0	-11.0	
9	1	RS		15.1	32.0	-2.0	
10	1	PA		25.63	36.0	-3.0	

4. Sort the data to get the following:

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

```
Select c.customer_state,  
Avg(i.freight_value) as Avg_freight_value,  
Dense_Rank() Over(Order by Avg(i.freight_value) desc) as Highest_Avg_Freight_Value,  
Dense_Rank() Over(Order by Avg(i.freight_value) asc) as Lowest_Avg_Freight_Value,  
from  
`target_retail.order_items` i join `target_retail.orders` o ON  
o.order_id = i.order_id join `target_retail.customers` c ON  
c.customer_id = o.customer_id  
Group by c.customer_state  
Order by Avg_freight_value desc  
Limit 5;
```

Query results

JOB INFORMATION		RESULTS	JSON	EXECUTION DETAILS	EXECUTION GRA
Row	customer_state	Avg_freight_value	Highest_Avg_Freight	Lowest_Avg_Freight	
1	RR	42.98442307692...	1	27	
2	PB	42.72380398671...	2	26	
3	RO	41.06971223021...	3	25	
4	AC	40.07336956521...	4	24	
5	PI	39.14797047970...	5	23	

Insights:

- RR, PB, RO, AC, PI are the top 5 states in Brazil, having the highest and lowest average freight value.

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

```
Select customer_state,
Avg(time_to_delivery) as Avg_time_to_delivery,
Dense_Rank() Over(Order by Avg(time_to_delivery) desc) as
Highest_Avg_time_to_delivery,
Dense_Rank() Over(Order by Avg(time_to_delivery) asc) as Lowest_Avg_time_to_delivery,
from
(
  Select c.customer_state,
  Date_diff(o.order_delivered_customer_date, o.order_purchase_timestamp, Day) as
time_to_delivery
  from `target_retail.orders` o inner join `target_retail.customers` c
  ON c.customer_id = o.customer_id
)
Group by customer_state
Order by Avg_time_to_delivery desc
Limit 5;
```

Query results

JOB INFORMATION		RESULTS	JSON	EXECUTION DETAILS	EXECUTION GRAPH	PREVIEW
Row	customer_state	Avg_time_to_delivery	Highest_Avg_time_to_delivery	Lowest_Avg_time_to_delivery		
1	RR	28.9756097560975...	1	27		
2	AP	26.7313432835820...	2	26		
3	AM	25.9862068965517...	3	25		
4	AL	24.0403022670025...	4	24		
5	PA	23.3160676532769...	5	23		

Insights:

- RR, AP, AM, AL, PA are the top 5 customer states with the highest/lowest average time to delivery.

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

```
Select customer_state,  
DENSE_RANK() Over(Order by diff_estimated_delivery) as slow_delivery,  
DENSE_RANK() Over(Order by diff_estimated_delivery desc) as fast_delivery  
from  
(  
Select c.customer_state,  
date_diff(o.order_estimated_delivery_date, o.order_delivered_customer_date, Day) as  
diff_estimated_delivery  
from `target_retail.orders` o inner join  
`target_retail.customers` c ON  
c.customer_id = o.customer_id  
)  
Group by customer_state, diff_estimated_delivery  
order by diff_estimated_delivery desc  
Limit 5;
```

Query results

JOB INFORMATION		RESULTS	JSON	EXECUTION DETAILS	
Row	customer_state		slow_delivery		fast_delivery
1	SP		198		1
2	MA		197		2
3	RS		196		3
4	SP		195		4
5	RJ		194		5

Insights:

- SP, MA, RS, SP, RJ are the top 5 states where delivery is fast and not so fast compared to estimated delivery.

Question 6:

Payment-type analysis:

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

```
Select
Extract(Month from o.order_purchase_timestamp)as Month,
Count(o.order_id) as order_count,
p.payment_type
from `target_retail.payments` p inner join
`target_retail.orders` o ON
o.order_id = p.order_id
Group by p.payment_type,Month
Order by Month, order_count desc;
```

Query results

JOB INFORMATION		RESULTS		JSON	EXECUTION DETAILS
Row	Month	order_count	payment_type		
1	1	6103	credit_c		payment_type
2	1	1715	UPI		
3	1	477	voucher		
4	1	118	debit_card		
5	2	6609	credit_card		
6	2	1723	UPI		
7	2	424	voucher		
8	2	82	debit_card		
9	3	7707	credit_card		
10	3	1942	UPI		

Insight:

- It is observed that the highest numbers of customers paid for their orders using credit card as the payment method.

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

```
Select payment_installments as No_of_installments,  
Count(order_id) as Order_count  
from `target_retail.payments`  
Group by payment_installments;
```

Query results

JOB INFORMATION		RESULTS	JSON
Row	No_of_installments	Order_count	
1	0	2	
2	1	52546	
3	2	12413	
4	3	10461	
5	4	7098	
6	5	5239	
7	6	3920	
8	7	1626	
9	8	4268	
10	9	644	

Insights:

- The majority of orders appear to be from installments 1 through 10, according to my observation.

Recommendations:

For the sake of business improvement and growth, I would like to suggest the following recommendations based on the conclusions I have taken from this case study:

- It is important to comprehend the causes of delivery delays to customers in various Brazilian locations. And, If something is related to the demographics or geographical situation of the regions, then, further investigation should be performed to improve the entire shipment pickup to delivery scenario.
- In order to reduce turmoil and the number of dissatisfied consumers, accurate predictions—or at the absolute least, the closest ones—should be made regarding the expected arrival of the orders. To make things simple and less hectic for customers, it should also be ensured that they receive adequate communication regarding the whereabouts of their shipments and the likelihood of them receiving the order.
- Keep track of customer evaluations concerning a variety of topics, such as product quality, delivery-related concerns, etc., to gain a thorough grasp of additional issues that consumers are experiencing but that the company has not yet noticed. To better understand consumer issues, it is crucial for the company to solicit client feedback at each stage between making a purchase and receiving the product.