

# Business Case: Target SQL

24 October 2022 19:27

## 1. Initial exploration of dataset like checking the characteristics of data

### 1. data type of columns in a table

<input type="checkbox"/>	Field name	Type	Mode	C
<input type="checkbox"/>	<a href="#">order_id</a>	STRING	NULLABLE	
<input type="checkbox"/>	<a href="#">customer_id</a>	STRING	NULLABLE	
<input type="checkbox"/>	<a href="#">order_status</a>	STRING	NULLABLE	
<input type="checkbox"/>	<a href="#">order_purchase_timestamp</a>	TIMESTAMP	NULLABLE	
<input type="checkbox"/>	<a href="#">order_approved_at</a>	TIMESTAMP	NULLABLE	
<input type="checkbox"/>	<a href="#">order_delivered_carrier_date</a>	TIMESTAMP	NULLABLE	
<input type="checkbox"/>	<a href="#">order_delivered_customer_date</a>	TIMESTAMP	NULLABLE	
<input type="checkbox"/>	<a href="#">order_estimated_delivery_date</a>	TIMESTAMP	NULLABLE	

### 2. Time period for which the data is given

```
1 select min(order_purchase_timestamp) as start_date, max(order_purchase_timestamp) as last_date,
2 date_diff(max(order_purchase_timestamp), min(order_purchase_timestamp), day) as time_period_in_days
3 from target.orders;
4
```

#### Query results

JOB INFORMATION					RESULTS	JSON	EXECUTION DETAILS	EXECUTION GRAPH	PREVIEW
Row	start_date	last_date	time_period_in_days						
1	2016-09-04 21:15:19 UTC	2018-10-17 17:30:18 UTC	772						

### 3. Cities and States covered in the dataset

```
1 select distinct c.customer_city as cities,
2 c.customer_state as states from target.orders o join target.customers c
3 on o.customer_id = c.customer_id
4 order by cities , states;
```

#### Query results

JOB INFORMATION					RESULTS	JSON	EXECUTION DETAILS	EXECUTION GRAPH
Row	cities	states						
1	abadia dos dourados	MG						
2	abadiania	GO						
3	abaete	MG						
4	abaetetuba	PA						
5	abaiara	CE						
6	abaira	BA						
7	abare	BA						
8	abatia	PR						
9	abdon batista	SC						
10	abelardo luz	SC						
11	abranes	BA						
12	abre campo	MG						
13	abreu e lima	PE						
14	acalaca	MG						
15	acallandia	MA						

## 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

```

6 select count(order_id) as number_of_orders,
7 extract( month from order_purchase_timestamp) as order_month,
8 extract(year from order_purchase_timestamp) as order_year
9 from target.orders
10 group by order_year, order_month
11 order by order_year;

```

### Query results

JOB INFORMATION		RESULTS	JSON	EXECUTION DETAILS
Row	number_of_...	order_month	order_year	
1	324	10	2016	
2	4	9	2016	
3	1	12	2016	
4	7544	11	2017	
5	5673	12	2017	
6	2404	4	2017	
7	4026	7	2017	
8	4631	10	2017	
9	3245	6	2017	
10	4285	9	2017	
11	1780	2	2017	
12	800	1	2017	
13	4331	8	2017	
14	2682	3	2017	
15	3700	5	2017	

Exploring the resultant data in excel, we see that there is a increase trend in number of orders year on year basis.

SUM of number_of_orders	order_year		
order_month	2016	2017	2018
1		800	7269
2		1780	6728
3		2682	7211
4		2404	6939
5		3700	6873
6		3245	6167
7		4026	6292
8		4331	6512
9	4	4285	16
10	324	4631	4
11		7544	
12	1	5673	
sum of orders	329	45101	54011

1	SUM of number_of_orders	order_year		
2	order_month	2016	2017	2018
3	1		800	7269
4	2		1780	6728
5	3		2682	7211
6	4		2404	6939
7	5		3700	6873
8	6		3245	6167
9	7		4026	6292
10	8		4331	6512
11	9	4	4285	16
12	10	324	4631	4
13	11		7544	
14	12	1	5673	
15	sum of orders	329	45101	54011
16				

Judging from the data available , we see highest peak of number of orders in month of October-November 2017 to January 2018 (towards er

2. What time do Brazilian customers tend to buy (Dawn, Morning, Afternoon or Night)?  
Mostly in afternoon and followed by night, morning and least at dawn respectively

```

37 select sum(orders) as num_of_orders, time_of_day from (
38   select count(*) as orders, order_purchase_timestamp,
39     extract(hour from order_purchase_timestamp) as hour_purchased,
40     CASE
41       WHEN extract(hour from order_purchase_timestamp) between 0 and 6 then 'dawn'
42       when extract(hour from order_purchase_timestamp) between 6 and 11 then 'morning'
43       when extract(hour from order_purchase_timestamp) between 12 and 17 then 'afternoon'
44       else 'night'
45     end
46     as time_of_day
47   from orders
48   where order_purchase_timestamp is not null
49   group by time_of_day, order_purchase_timestamp
50   order by orders
51 )x
52 group by x.time_of_day
53

```

00% 23:52 4 errors found

Result Grid Filter Rows: Search Export:

num_of_orders	time_of_day
21738	morning
38361	afternoon
34100	night
5242	dawn

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

1. Get month on month orders by region, states

```

54 select customer_state, sum(num_of_orders) as total_orders, order_year, order_month from (
55   select c.customer_state, extract( month from o.order_purchase_timestamp) as order_month,
56     extract(year from o.order_purchase_timestamp) as order_year,
57     count(*) as num_of_orders
58   from orders o left join customers c
59   on o.customer_id = c.customer_id
60   group by c.customer_state, order_year, order_month, o.order_id
61 )x
62 group by order_year, order_month, customer_state
63 order by order_year, order_month, customer_state

```

100% 1:64 4 errors found

Result Grid Filter Rows: Search Export:

customer_state	total_orders	order_year	order_month
RR	1	2016	9
RS	1	2016	9
SP	2	2016	9
AL	2	2016	10
BA	4	2016	10
CE	8	2016	10
DF	6	2016	10
ES	4	2016	10
GO	9	2016	10
MA	4	2016	10
MG	40	2016	10
MT	3	2016	10
PA	4	2016	10
PB	1	2016	10
PE	7	2016	10
PI	1	2016	10
PR	19	2016	10
RJ	56	2016	10
RN	4	2016	10
RR	1	2016	10
RS	24	2016	10
SC	11	2016	10
SE	3	2016	10

2. How are customers distributed in Brazil?

State wise distribution of customers in Brazil:

```

67 select count(*) as total_customers, customer_state
68 from customers
69 group by customer_state
70 order by customer_state
71

```

100% 24:70 4 errors found

Result Grid Filter Rows: Search Export:

total_custom...	customer_state
81	AC
413	AL
149	AM
68	AP
3360	BA
1336	CE
2140	DF
2033	ES
2020	GO
747	MA
11635	MG
715	MS
907	MT
975	PA
536	PB
1652	PE
495	PI
5045	PR
12852	RJ
485	RN
253	RO
48	RR
5466	RS
3637	SC
350	SE
41746	SP
280	TO

4. Impact on Economy: Analyze the money movemented 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)

```

82 select total_price, order_year, total_price_2017, round(((total_price - total_price_2017)/total_price)*100,2) as per_price_increase from (
83 select round(sum(o1.price),2) as total_price, extract( year from o.order_purchase_timestamp) as order_year,
84 lag(round(sum(o1.price),2),1) over( order by extract( year from o.order_purchase_timestamp)) as total_price_2017
85 from orders o left join order_items o1
86 on o.order_id=o1.order_id
87 where extract( year from o.order_purchase_timestamp) between 2017 and 2018 and extract( month from o.order_purchase_timestamp) between 1 and 8
88 group by order_year
89 ) x
90 where x.total_price_2017 is not null
91

```

100% 181 4 errors found

Result Grid Filter Rows: Search Export:

total_price	order_year	total_price_2017	per_price_increase
7288906.9	2018	3113000.32	57.85

[illegible]

1. Calculate days between purchasing, c

---

```
109 select extract( year from order_purchase_timestamp) as order_year,
110 #datediff(order_delivered_customer_date,order_purchase_timestamp)as "diff(purchase,delivered)",
```

---

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

```

119 alter table orders
120 add column time_to_delivery int;
121
122 UPDATE orders SET time_to_delivery=datediff(order_delivered_customer_date,order_purchase_timestamp) where order_delivered_customer_date <> '' and
123 order_purchase_timestamp <> '';
124
125 alter table orders
126 add column diff_estimated_delivery int;
127
128 UPDATE orders SET diff_estimated_delivery=datediff(order_estimated_delivery_date,order_delivered_customer_date) where order_delivered_customer_date <> '' and
129 order_estimated_delivery_date <> '';
130
131 select order_purchase_timestamp, order_delivered_customer_date,
132 order_estimated_delivery_date,time_to_delivery, diff_estimated_delivery
133 from orders
134

```

order_purchase_timestamp	order_delivered_customer_date	order_estimated_delivery_date	time_to_delivery	diff_estimated_delivery
2017-10-09 10:56:33	2017-10-10 21:26:13	2017-10-18 00:00:00	8	8
2018-07-24 20:41:37	2018-08-07 15:27:45	2018-08-13 00:00:00	14	6
2018-08-08 08:18:44	2018-08-17 18:04:29	2018-09-04 00:00:00	9	18
2017-11-18 19:28:06	2017-12-02 00:38:42	2017-12-15 00:00:00	14	15
2018-03-13 21:18:39	2018-03-16 18:17:02	2018-03-24 00:00:00	3	10
2017-07-09 21:57:05	2017-07-26 10:57:55	2017-08-21 00:00:00	17	6
2017-04-11 12:22:58		2017-05-09 00:00:00	29	29
2017-06-16 18:10:20	2017-06-26 12:55:51	2017-07-07 00:00:00	10	12
2017-01-23 18:33:29	2017-02-03 14:08:10	2017-03-03 00:00:00	10	32
2017-07-29 11:55:52	2017-08-16 17:14:30	2017-08-23 00:00:00	18	7
2017-05-18 15:41:10	2017-05-25 11:15:31	2017-06-07 00:00:00	13	9
2017-07-13 19:58:11	2017-07-19 14:04:48	2017-08-08 00:00:00	8	20
2018-06-07 00:08:19	2018-06-19 12:05:52	2018-07-18 00:00:00	12	29
2018-07-25 17:44:10	2018-07-30 15:52:25	2018-08-08 00:00:00	5	9
2018-05-01 14:14:08	2018-05-15 23:16:56	2018-05-21 00:00:00	11	16

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

```

139 select c.customer_state,
140 round(avg(oi.freight_value),2) as avg_freight,
141 round(avg(o.time_to_delivery),2) as avg_time_to_delivery,
142 round(avg(o.diff_estimated_delivery),2) as diff_estimated_delivery
143 from orders o left join order_items oi on o.order_id = oi.order_id
144 join customers c on o.customer_id = c.customer_id
145 group by c.customer_state
146 order by c.customer_state
147
148

```

customer_state	avg_freight	avg_time_to_delivery	diff_estimated_delivery
AC	40.07	20.88	20.98
AL	35.84	24.45	6.74
AM	33.21	26.34	19.93
AP	34.01	28.22	18.40
BA	26.36	19.19	10.98
CE	32.71	20.92	11.10
DF	21.04	12.89	12.20
ES	22.06	15.59	10.65
GO	22.77	15.34	12.29
MA	38.26	21.59	9.91
MG	20.63	11.92	13.34
MS	23.37	15.46	11.23
MT	28.17	17.91	14.57
PA	35.83	23.70	14.25
PB	42.72	20.55	13.04
PE	32.92	18.22	13.45
PI	39.15	19.32	11.53
PR	29.63	11.89	13.49

4. Sort the data to get the following:
- Top 5 states with highest/lowest average freight value - sort in desc/asc limit 5

5 highest:

```

149 select c.customer_state, oi.freight_value
150 from orders o left join order_items oi on o.order_id = oi.order_id
151 join customers c on o.customer_id = c.customer_id
152 group by customer_state, freight_value
153 order by freight_value desc
154 limit 5
155
156

```

customer_state	freight_value
PI	409.68
SC	375.28
PR	375.28
SP	339.59
MT	338.3

5 lowest:

```

149 select c.customer_state, oi.freight_value
150 from orders o left join order_items oi on o.order_id = oi.order_id
151 join customers c on o.customer_id = c.customer_id
152 where oi.freight_value is not null
153 group by customer_state, freight_value
154 order by freight_value
155 limit 5
156
157

```

customer_state	freight_value
PA	0
RN	0
TO	0
MT	0

MA	0
----	---

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

Top 5 state with lowest avg time to deliver:

```

149 select c.customer_state, avg(o.time_to_delivery)
150 from orders o left join order_items oi on o.order_id = oi.order_id
151 join customers c on o.customer_id = c.customer_id
152 group by customer_state
153 order by avg(o.time_to_delivery)
154 limit 5
155
156

```

100% 2:155 5 errors found

Result Grid Filter Rows: Search Export: Fetch rows:

customer_state	avg(o.time_to_delive...
SP	8.6623
PR	11.8831
MG	11.9207
DF	12.8838
SC	14.9502

Top 5 states with highest time\_to\_deliver:

```

149 select c.customer_state, avg(o.time_to_delivery)
150 from orders o left join order_items oi on o.order_id = oi.order_id
151 join customers c on o.customer_id = c.customer_id
152 group by customer_state
153 order by avg(o.time_to_delivery) desc
154 limit 5
155
156

```

100% 8:154 5 errors found

Result Grid Filter Rows: Search Export: Fetch rows:

customer_state	avg(o.time_to_delive...
AP	28.2222
RR	28.1739
AM	28.3374
AL	24.4473
PA	23.7021

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

5 states with really fast delivery:(on an avg over the years)

```

156 select c.customer_state, avg(o.diff_estimated_delivery)
157 from orders o left join order_items oi on o.order_id = oi.order_id
158 join customers c on o.customer_id = c.customer_id
159 group by customer_state #,diff_estimated_delivery
160 order by avg(diff_estimated_delivery)
161 limit 5
162
163

```

100% 8:161 5 errors found

Result Grid Filter Rows: Search Export: Fetch rows:

customer_state	avg(o.diff_estimated_delivery)
AL	8.7354
MA	9.9063
SE	10.0027
ES	10.6463
BA	10.9826

5 states with not so fast delivery: (on an avg over the years)

```

156 select c.customer_state, avg(o.diff_estimated_delivery)
157 from orders o left join order_items oi on o.order_id = oi.order_id
158 join customers c on o.customer_id = c.customer_id
159 group by customer_state #,diff_estimated_delivery
160 order by avg(diff_estimated_delivery) desc
161 limit 5
162
163

```

100% 8:161 5 errors found

Result Grid Filter Rows: Search Export: Fetch rows:

customer_state	avg(o.diff_estimated_delivery)
AC	20.9780
RO	20.0403
AM	19.9325
AP	18.3951
RR	18.3261

We can take a look at the logistics on the top 5 states with fastest delivery time on an average and can implement similar optimization c Portray the estimated delivery date as per actual delivery in the states with not so fast delivery.

## 6. Payment type analysis:

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

```

163 select p.payment_type, extract( year from order_purchase_timestamp) as order_year,
164 extract(month from o.order_purchase_timestamp) as order_month, count(*) as num_of_orders
165 from orders o left join payments p on o.order_id=p.order_id
166 where payment_type is not null
167 group by payment_type,order_year, order_month
168 order by order_year,order_month

```

169 100% 33:168 5 errors found

Result Grid Filter Rows: Search Export:

payment_type	order_year	order_month	num_of_orders
credit_card	2016	9	3
debit_card	2016	10	2
voucher	2016	10	23
UPI	2016	10	63
credit_card	2016	10	254
credit_card	2016	12	1
voucher	2017	1	61
debit_card	2017	1	9
credit_card	2017	1	583
UPI	2017	1	197
credit_card	2017	2	1356
UPI	2017	2	388
debit_card	2017	2	13
voucher	2017	2	119
voucher	2017	3	200
credit_card	2017	3	2016
UPI	2017	3	599
debit_card	2017	3	31
credit_card	2017	4	1846
voucher	2017	4	202
debit_card	2017	4	27
UPI	2017	4	496

## 2. Distribution of payment instalments and count of orders

```

170 select payment_installments, count(*) as count_orders
171 from payments
172 group by payment_installments
173 order by payment_installments
174

```

100% 30:173 5 errors found

Result Grid Filter Rows: Search Export:

payment_installme...	count_orders
0	2
1	52546
2	12413
3	10461
4	7098
5	5239
6	3920
7	1626
8	4268
9	644
10	5328
11	23
12	133
13	16
14	15
15	74
16	5
17	8
18	27
20	17
21	3
22	1

### Actionable Insights:

- The most orders comes around year end and new year. Target can scale up their infrastructure specially during this time and can sales even further
- Few states have considerable delta between estimated delivery date and actual delivery date. We can work with our logistics par network and reduce the time between order purchase and customer delivery. We can also tweak/optimize our estimated delivery delivery dates more accurately
- We see that there are many orders purchased via credit cards and we can partner with leading credit card provider in the country made via credit cards to even boost sales

### Recommendations:

- To work with our delivery partners to reduce the time delta between purchase date and customer delivery date.
- For the states with highest freight charges , we can find alternate routes or work with freight institutions for exclusive partnerships time of delivery to customers
- Since there was a high YoY percentage of cost increase, we need to find ways to bring the costs down for a better sustainable bu on customer experience

