SQL BUSINESS CASE – “TARGET”

We have been provided with the data of one of the leading retailers in America, “TARGET”. These datasets contain information on 100k orders from 2016 to 2018 made at Target in Brazil. So, we can view this information available from multiple dimensions and analyze the given data to come up with some meaningful insight and recommendations which can benefit Target in their business.

Exploratory Analysis:

1) What is the data type of the data available in the datasets?

We can use the below query to know the data types available in the datasets.

```
SELECT
  COLUMN_NAME, DATA_TYPE
FROM
  Target.INFORMATION_SCHEMA.COLUMNS
WHERE
  TABLE_SCHEMA = 'Target'
AND
  TABLE_NAME = 'customers';
```

The above query yields the below result for the customer table. We can see the customer table has all string variables and one variable customer_zip_code_prefix is an integer.

Customers:

COLUMN_NAME	DATA_TYPE
customer_id	STRING
customer_unique_id	STRING
customer_zip_code_prefix	INT64
customer_city	STRING
customer_state	STRING

We can run the same query for all the other datasets/tables(replacing the table name in the above query) and see the data type.

Geolocation:

COLUMN_NAME	DATA_TYPE
geolocation_zip_code_prefix	INT64
geolocation_lat	FLOAT64
geolocation_lng	FLOAT64
geolocation_city	STRING
geolocation_state	STRING

Orders:

order_id	STRING
customer_id	STRING
order_status	STRING
order_purchase_timestamp	TIMESTAMP
order_approved_at	TIMESTAMP
order_delivered_carrier_date	TIMESTAMP
order_delivered_customer_date	TIMESTAMP
order_estimated_delivery_date	TIMESTAMP

Since the orders table is supposed to have details of delivery details, we have a new data type as TIMESTAMP for some of the variables.

Order_items:

COLUMN_NAME	DATA_TYPE
order_id	STRING
order_item_id	INT64
product_id	STRING
seller_id	STRING
shipping_limit_date	TIMESTAMP
price	FLOAT64
freight_value	FLOAT64

Order_reviews:

COLUMN_NAME	DATA_TYPE
review_id	STRING
order_id	STRING
review_score	INT64
review_comment_title	STRING
review_creation_date	TIMESTAMP
review_answer_timestamp	TIMESTAMP

Payments:

COLUMN_NAME	DATA_TYPE
order_id	STRING
payment_sequential	INT64
payment_type	STRING
payment_installments	INT64
payment_value	FLOAT64

Products:

COLUMN_NAME	DATA_TYPE
product_id	STRING
product_category	STRING
product_name_length	INT64
product_description_length	INT64
product_photos_qty	INT64
product_weight_g	INT64
product_length_cm	INT64
product_height_cm	INT64
product_width_cm	INT64

Sellers:

seller_id	STRING
seller_zip_code_prefix	INT64
seller_city	STRING
seller_state	STRING

2. Time period of the given data:

The 1st order purchase date is of 4th September 2016 and the latest purchase date is of 17th October 2018. So, the orders table contains orders for the period of 773 days starting from 4th September 2016.

```
select min(date(a.order_purchase_timestamp)) from Target.orders as A;
select max(date(a.order_purchase_timestamp)) from Target.orders as A;

select (max(date(a.order_purchase_timestamp)) -
min(date(a.order_purchase_timestamp))) as duration from  Target.orders as A;
```

3. Cities and states covered in the dataset:

If we take the customers and orders table, we can see there are 4310 distinct states and cities are covered.

```
select distinct customer_city, customer_state from Target.customers A,
Target.orders B
where A.customer_id = B.customer_id ;
```

In-depth Exploration

4. How the business has increased/decreased over the entire duration?

```
select year, count(order_id) as orders_count from
(
select *, EXTRACT(year FROM order_purchase_timestamp) as year from Target.orders
)
group by year
order by year;
```

year	orders_count
2016	329
2017	45101
2018	54011

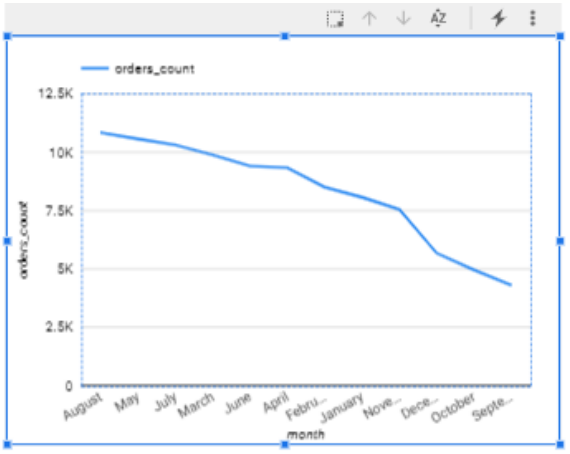
We can see there is a drastic increase in the number of orders in 2017 and it has increased significantly in the year 2018 as well.

```sql
select year,month, count(order_id) as orders_count from
(
select *, EXTRACT(year FROM order_purchase_timestamp) as year , EXTRACT(month FROM
order_purchase_timestamp) as month from Target.orders
)
group by year, month
order by year, month;
```

year	month	orders_count
2016	9	4
2016	10	324
2016	12	1
2017	1	800
2017	2	1780
2017	3	2682
2017	4	2404
2017	5	3700
2017	6	3245
2017	7	4026
2017	8	4331
2017	9	4285
2017	10	4631
2017	11	7544
2017	12	5673
2018	1	7269
2018	2	6728
2018	3	7211
2018	4	6939
2018	5	6873
2018	6	6167
2018	7	6292
2018	8	6512
2018	9	16
2018	10	4

	month	orders_count
1.	August	10,843
2.	May	10,573
3.	July	10,318
4.	March	9,893
5.	June	9,412
6.	April	9,343
7.	February	8,508
8.	January	8,069
9.	November	7,544
10.	December	5,674
11.	October	4,959

1 - 12 / 12



We can clearly see a peak in sales in the months of August, May, and July. Also, in the month of November and December in 2017, we have a peak in sales which probably be due to the festive season.

5. Do we have any trend in buying as per the time of the day?

```
SELECT CASE WHEN EXTRACT(HOUR FROM o.order_purchase_timestamp)>=4 and
EXTRACT(HOUR FROM o.order_purchase_timestamp)<=6 THEN 'DAWN'
WHEN EXTRACT(HOUR FROM o.order_purchase_timestamp)>6 AND
EXTRACT(HOUR FROM o.order_purchase_timestamp)<12 THEN 'MORNING'
WHEN EXTRACT(HOUR FROM o.order_purchase_timestamp)>=12 AND
EXTRACT(HOUR FROM o.order_purchase_timestamp)<=18 THEN 'AFTERNOON'
ELSE 'NIGHT'
END AS TIME_OF_DAY, count(distinct(o.order_id)) as orders_placed
FROM Target.orders o GROUP BY TIME_OF_DAY order by orders_placed desc ;
```

TIME_OF_DAY	orders_placed
AFTERNOON	44130
NIGHT	32677
MORNING	21738
DAWN	896

6. Which date of the month had the highest sales:

```
select year , month, max(day_highest_no_of_orders) day_highest_no_of_orders,
max(max_orders) orders from
(select year, month, nth_value(day,1) over (partition by year, month order by
orders_placed desc range between unbounded preceding and unbounded following ) as
```

```
day_highest_no_of_orders, max(orders_placed) over (partition by year, month) as
max_orders from
(
select EXTRACT(year FROM o.order_purchase_timestamp) as year,EXTRACT(month FROM
o.order_purchase_timestamp) as month,EXTRACT(day FROM o.order_purchase_timestamp)
as day, count(o.order_id) as orders_placed
FROM Target.orders o GROUP BY year, month, day order by year asc, month asc,
orders_placed desc
)

order by year, month)
group by year, month
order by year, month
;
```

year	month	day_highest_no_of_orders	orders
2017	1	26	86
2017	2	7	112
2017	3	20	119
2017	4	26	125
2017	5	29	160
2017	6	19	156
2017	7	18	192
2017	8	15	194
2017	9	13	207
2017	10	17	202
2017	11	24	1176
2017	12	4	337
2018	1	22	314
2018	2	28	313
2018	3	19	303
2018	4	19	293
2018	5	7	372
2018	6	11	294
2018	7	31	322
2018	8	6	372
2018	9	3	4
2018	10	1	1

Evolution of E-commerce orders in the Brazil region:

7. States with the highest number of orders month on month:

```
select distinct year, month, nth_value(customer_state,1) over (partition by year,
month order by cnt desc range between unbounded preceding and unbounded following)
as highest_orders_state from (

select customer_state, EXTRACT(year FROM o.order_purchase_timestamp) year,
EXTRACT(month FROM o.order_purchase_timestamp) month, count(order_id) as cnt
from Target.customers c, Target.orders o
where o.customer_id=c.customer_id
group by customer_state, year , month
order by year, cnt desc)
;
```

year	month	highest_orders_state
2017	1	SP
2017	2	SP
2017	3	SP
2017	4	SP
2017	5	SP
2017	6	SP
2017	7	SP
2017	8	SP
2017	9	SP
2017	10	SP
2017	11	SP
2017	12	SP
2018	1	SP
2018	2	SP
2018	3	SP
2018	4	SP
2018	5	SP
2018	6	SP
2018	7	SP
2018	8	SP

8. Which region/postal code has the highest number of sales month by month:

```
select distinct year, month, nth_value(customer_zip_code_prefix,1) over (partition by year, month order by cnt desc range between unbounded preceding and unbounded following) as highest_order_region from (

select customer_zip_code_prefix, EXTRACT(year FROM o.order_purchase_timestamp) year, EXTRACT(month FROM o.order_purchase_timestamp) month, count(order_id) as cnt
from Target.customers c, Target.orders o
where o.customer_id=c.customer_id
group by customer_zip_code_prefix, year, month
order by year, cnt desc)
;
```

year	month	highest_order_region
2017	1	80030
2017	2	22775
2017	3	35500
2017	4	13280
2017	5	22790
2017	6	42700
2017	7	37550
2017	8	13087
2017	9	20520
2017	10	36570
2017	11	22790
2017	12	22790
2018	1	22793
2018	2	22793
2018	3	24220
2018	4	24230
2018	5	30140
2018	6	38400
2018	7	24210
2018	8	13212

9. The region with the highest number of orders:

```
select customer_zip_code_prefix, EXTRACT(month FROM o.order_purchase_timestamp)
month, count(order_id) as cnt
from Target.customers c, Target.orders o
where o.customer_id=c.customer_id
group by customer_zip_code_prefix, month
order by month, cnt desc ;
```

customer_zip_code_p...	cnt
22790	142
24220	124
22793	121
24230	117
22775	110
29101	101
13212	95
35162	93
22631	89
38400	87

10. Region/zip code with the highest number of customers:

```
select customer_zip_code_prefix, count(customer_id) as count_of_customers
from Target.customers
group by customer_zip_code_prefix
order by count_of_customers desc limit 10;
```

customer_zip_code_prefix	count_of_customers
22790	142
24220	124
22793	121
24230	117
22775	110
29101	101
13212	95
35162	93
22631	89
38400	87

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

11. How Business expanded in terms of money for Jan to August months of the year.

```
select *,((lead_order_amount - total_order_cost)/ total_order_cost)*100 as
percent_inc from (
select *, lead(total_order_cost) over (order by total_order_cost ) as
lead_order_amount from
(SELECT extract(year from o.order_purchase_timestamp) year,
ROUND(SUM(oi.price+oi.freight_value),2) as total_order_cost
from Target.order_items oi join Target.orders o
on oi.order_id=o.order_id

WHERE EXTRACT(MONTH FROM o.order_purchase_timestamp) BETWEEN 1 AND 8
group by year
order by year));
```

year	total_order_cost	lead_order_amount	percent_inc
2018	8643531.14	*null*	*null*
2017	3610270.15	8643531.14	139.415079...

There is a whopping increase of 139% for total business done in terms of money from 2017 to 2018.

12. Top 5 states with the highest average freight value?

```
select c.customer_state,round(AVG(oi.freight_value),2) as AVG_FREIGHT_VALUE
from Target.customers c join Target.orders o
on c.customer_id=o.customer_id
join Target.order_items oi
on o.order_id=oi.order_id
GROUP BY c.customer_state
order by AVG_FREIGHT_VALUE DESC LIMIT 5;
```

customer_state	AVG_FREIGHT_VALUE
RR	42.98
PB	42.72
RO	41.07
AC	40.07
PI	39.15

13. Bottom 5 states with the lowest avg freight value:

```
select c.customer_state,round(AVG(oi.freight_value),2) as AVG_FREIGHT_VALUE
from Target.customers c join Target.orders o
on c.customer_id=o.customer_id
join Target.order_items oi
on o.order_id=oi.order_id
GROUP BY c.customer_state
order by AVG_FREIGHT_VALUE asc LIMIT 5;
```

customer_state	AVG_FREIGHT_VALUE
SP	15.15
PR	20.53
MG	20.63
RJ	20.96
DF	21.04

14. Bottom 5 states with the lowest total freight value:

```
select c.customer_state,round(sum(oi.freight_value),2) as total_freight_value
from Target.customers c join Target.orders o
on c.customer_id=o.customer_id
join Target.order_items oi
```

```
on o.order_id=oi.order_id
GROUP BY c.customer_state
order by total_freight_value asc LIMIT 5;
```

customer_state	total_freight...
RR	2235.19
AP	2788.5
AC	3686.75
AM	5478.89
RO	11417.38

15. Top 5 states with the highest total freight value:

```
select c.customer_state,round(sum(oi.freight_value),2) as total_freight_value
from Target.customers c join Target.orders o
on c.customer_id=o.customer_id
join Target.order_items oi
on o.order_id=oi.order_id
GROUP BY c.customer_state
order by total_freight_value desc LIMIT 5;
```

customer_state	total_freight_value
SP	718723.07
RJ	305589.31
MG	270853.46
RS	135522.74
PR	117851.68

From the above results on avg and total freight values, we can see the relationship between freight value and sales. The lower freight value enhances the chance for more selling of the products.

Analysis on sales, freight, and delivery time

16. 5 states with the highest average delivery time:

```
Select c.customer_state,
round(avg(date_diff(order_delivered_customer_date,order_purchase_timestamp,
day)),2)
```

```
as avg_delivery_time
from Target.customers c join Target.orders o
on c.customer_id=o.customer_id
group by c.customer_state
order by avg_delivery_time desc limit 5;
```

customer_state	avg_delivery_time
RR	28.98
AP	26.73
AM	25.99
AL	24.04
PA	23.32

17. 5 states where delivery is speedy:

```
select
c.customer_state,round(avg(date_diff(order_estimated_delivery_date,order_delivered_
customer_date, day)),2) as act_est_del_diff
from Target.customers c join Target.orders o
on c.customer_id=o.customer_id
group by c.customer_state
order by act_est_del_diff desc limit 5;
```

customer_state	act_est_del_diff
AC	19.76
RO	19.13
AP	18.73
AM	18.61
RR	16.41

18. 5 states where delivery is slow:

```
select
c.customer_state,round(avg(date_diff(order_estimated_delivery_date,order_delivered_
customer_date, day)),2) as act_est_del_diff
from Target.customers c join Target.orders o
on c.customer_id=o.customer_id
```

```
group by c.customer_state
order by act_est_del_diff asc limit 5;
```

customer_state	act_est_del_diff
AL	7.95
MA	8.77
SE	9.17
ES	9.62
BA	9.93

19. The average difference between purchase and estimated delivery time:

```
SELECT
round(avg(date_diff(order_estimated_delivery_date,order_purchase_timestamp,day)),2)
as avg_estimated_delivery_time
from Target.orders;
```

avg_estimated_delivery_time_in_days
23.4

20. Top 10 delayed estimated delivery times:

```
SELECT date_diff(order_estimated_delivery_date,order_purchase_timestamp,day) as
estimated_delivery_time_in_days
from Target.orders
order by estimated_delivery_time_in_days desc limit 10;
```

estimated_delivery_time_in_days
155
149
146
144
144
142
140
116
109
106

21. Correlation between review score and estimated delivery time:

```
SELECT review_score, EXTRACT(year FROM order_purchase_timestamp) as year,
round(avg(date_diff(order_estimated_delivery_date,order_purchase_timestamp,day)),2)
as avg_estimated_delivery_time_in_days
from Target.orders O, Target.order_reviews R
where O.order_id=R.order_id
group by year, review_score
order by year, review_score;
```

review_score	year	avg_estimated_delivery_time_in_days
1	2016	52.25
2	2016	52.0
3	2016	54.5
4	2016	55.16
5	2016	55.72
1	2017	25.14
2	2017	25.13
3	2017	24.63
4	2017	24.53
5	2017	24.07

1	2018	23.31
2	2018	23.27
3	2018	23.2
4	2018	22.73
5	2018	21.99

From the above table, we can see the review score does not seems to be much affected by estimated delivery time. Also, we can notice the average estimated delivery time has improved a lot with time.

22. Now, we can check if the review score has any relation to early delivery than the estimated delivery.

```
SELECT review_score, EXTRACT(year FROM order_purchase_timestamp) as
year,round(avg(date_diff(order_estimated_delivery_date,order_delivered_customer_dat
e,day)),2) as diff_actual_est_delivery, count(review_score) as cnt
from Target.orders O, Target.order_reviews R
where O.order_id=R.order_id
group by year, review_score
order by year, review_score;
```

review_score	year	diff_actual_est_delivery	cnt
1	2016	27.38	89
2	2016	34.17	8
3	2016	34.62	22
4	2016	34.76	51
5	2016	38.47	156
1	2017	4.29	5049
2	2017	8.51	1466
3	2017	10.43	3815
4	2017	11.67	8889
5	2017	12.78	25825
1	2018	2.65	6286
2	2018	7.17	1677
3	2018	9.29	4342
4	2018	11.11	10202
5	2018	11.99	31347

We can observe here with earlier delivery we have a better review score. So, earlier delivery can be a factor in a better review score.

23. Total cost of all the orders for the entire duration:

```
select round(sum(freight_value+price),2) as total_cost from Target.order_items;
```

total_cost
15843553.24

24. Top 10 orders in terms of cost across the entire duration:

```
select order_id, round(sum(freight_value+price),2) total_cost from
Target.order_items
group by order_id
order by total_cost desc limit 10;
```

order_id	total_cost
03caa2c082116e1d31e67e9ae...	13664.08
736e1922ae60d0d6a89247b8...	7274.88
0812eb902a67711a1cb742b3c...	6929.31
fefacc66af859508bf1a7934ea...	6922.21
f5136e38d1a14a4dbd87dff67d...	6726.66
2cc9089445046817a7539d90...	6081.54
a96610ab360d42a2e5335a39...	4950.34
b4c4b76c642808cbe472a32b8...	4809.44
199af31afc78c699f0dbf71fb1...	4764.34
8dbc85d1447242f3b127dda39...	4681.78

Payment Type Analysis:

25. Month over month order count for various payment methods:

```
SELECT distinct(p.payment_type) as payment_methods, extract(month from
o.order_purchase_timestamp) as month,
```

```
count(distinct(p.order_id)) as order_count from Target.orders o join
Target.payments p
on o.order_id=p.order_id
group by Month,payment_methods
order by order_count desc;
```

payment_methods	month	order_count
credit_card	5	8308
credit_card	8	8235
credit_card	7	7810
credit_card	3	7682
credit_card	4	7276
credit_card	6	7248
credit_card	2	6582
credit_card	1	6093
credit_card	11	5867
credit_card	12	4364
credit_card	10	3763
credit_card	9	3277
UPI	8	2077
UPI	7	2074
UPI	5	2035
UPI	3	1942
UPI	6	1807
UPI	4	1783
UPI	2	1723
UPI	1	1715
UPI	11	1509
UPI	12	1160
UPI	10	1056
UPI	9	903
voucher	8	430
voucher	7	417
voucher	3	395
voucher	5	374
voucher	6	373
voucher	4	353

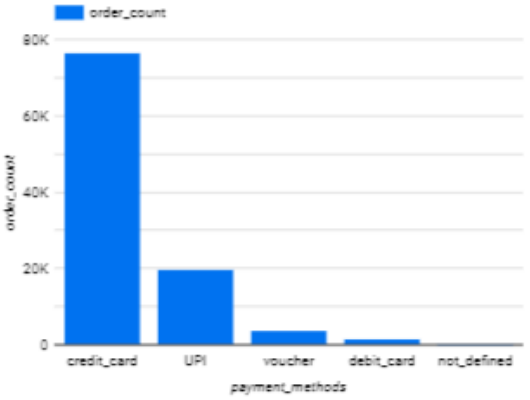
voucher	1	337
debit_card	8	311
voucher	2	288
voucher	11	267
debit_card	7	264
voucher	10	223
voucher	12	220
debit_card	6	208
voucher	9	189
debit_card	4	124
debit_card	1	118
debit_card	3	109
debit_card	2	82
debit_card	5	81
debit_card	11	70
debit_card	12	64
debit_card	10	54
debit_card	9	43
not_defined	8	2
not_defined	9	1

26. Order count payment method wise:

```
select distinct(payment_type) as payment_methods, count(distinct(order_id)) as no_of_orders from Target.payments group by payment_methods order by no_of_orders desc;
```

	payment_methods	order_count
1.	credit_card	76,505
2.	UPI	19,784
3.	voucher	3,866
4.	debit_card	1,528
5.	not_defined	3

1 - 5 / 5



Further exploration on the data

27. We can also do some exploratory analysis from the other datasets. E.g. How many sellers are registered with TARGET?

```
SELECT COUNT(DISTINCT(seller_id)) as NO_OF_SELLERS FROM Target.sellers;
```

We have 3095 sellers registered with Target.

NO_OF_SELLERS
3095

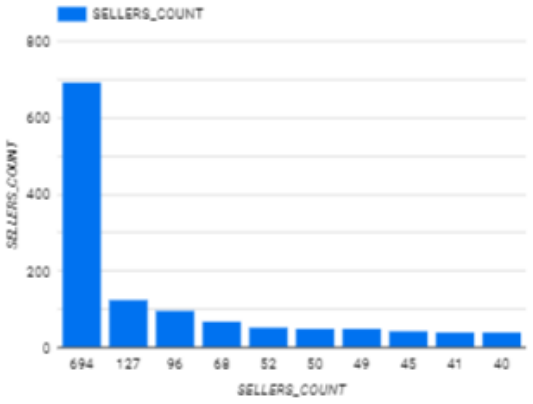
28. Top 10 and Bottom 10 cities for sellers registered with Target.

```
SELECT seller_city, COUNT(DISTINCT(seller_id)) as SELLERS_COUNT FROM Target.sellers
GROUP BY seller_city ORDER BY SELLERS_COUNT DESC LIMIT 10;
```

seller_city	SELLERS_COUNT
sao paulo	694
curitiba	127
rio de janeiro	96
belo horizonte	68
ribeirao preto	52
guarulhos	50
ibitinga	49
santo andre	45
campinas	41
maringa	40

	seller_city	SELLERS_COUNT
1.	sao paulo	694
2.	curitiba	127
3.	rio de janeiro	96
4.	belo horizonte	68
5.	ribeirao preto	52
6.	guarulhos	50
7.	ibitinga	49
8.	santo andre	45
9.	campinas	41
10.	maringa	40

1 - 10 / 10




```
SELECT seller_city, COUNT(DISTINCT(seller_id)) as SELLERS_COUNT FROM Target.sellers
GROUP BY seller_city ORDER BY SELLERS_COUNT asc LIMIT 10;
```

seller_city	SELLERS_COUNT
ilheus	1
eunapolis	1
feira de santana	1
porto seguro	1
manaus	1
rio branco	1
barro alto	1
guanambi	1
irece	1
bahia	1

We can check the count of sellers grouped basis states and cities together and check the top count of the sellers.

29. Top 10 unique combinations of states and cities with the highest number of sellers:

```
select seller_state, seller_city, count(distinct seller_id) as no_of_sellers from
Target.sellers
group by seller_state, seller_city
order by no_of_sellers desc limit 10;
```

seller_state	seller_city	no_of_sellers
SP	sao paulo	694
PR	curitiba	124
RJ	rio de janeiro	93
MG	belo horizonte	66
SP	ribeirao preto	52
SP	guarulhos	50
SP	ibitinga	49
SP	santo andre	45
SP	campinas	41
PR	maringa	40

We can notice the count is similar to the count obtained basis top seller grouped by City. However, few cities here have a little less count of sellers when compared to the count obtained only on the basis of the city.
So, it seems like there are a few cities that are tagged to more than one state. There may be some inaccuracy in the demographic data in the seller's table.

```
select seller_city, count(distinct seller_state) as multstate from (select
seller_state, seller_city, count(distinct seller_id) as no_of_sellers from
Target.sellers
group by seller_state, seller_city)
group by seller_city
having multstate > 1;
```

Sample of some of the cities tagged to multiple states.

seller_city	multstate
ipira	2
vila velha	2
andradas	2
jacutinga	2
juiz de fora	2
belo horizonte	2
marechal candido rondon	3
guaira	2

There are 23 such cities tagged to multiple states.

30. Top 5 states contributing to the highest number of sellers registration:

```
SELECT seller_state, COUNT(DISTINCT(seller_id)) as NO_OF_SELLERS FROM
Target.sellers
GROUP BY seller_state ORDER BY NO_OF_SELLERS DESC LIMIT 5;
```

seller_state	NO_OF_SELLERS
SP	1849
PR	349
MG	244
SC	190
RJ	171

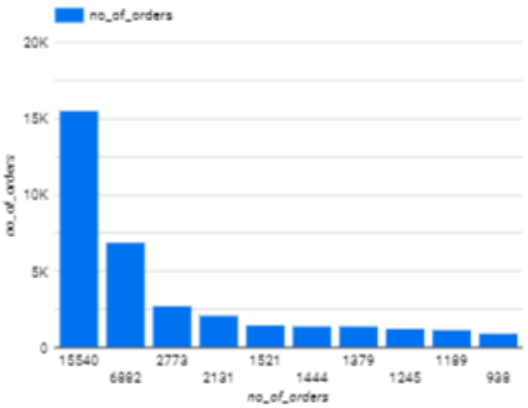
31. Top 10 cities with the highest orders:

```
select c.customer_city, count(distinct(order_id)) as no_of_orders from
Target.customers c, Target.orders o
where c.customer_id=o.customer_id
group by customer_city order by no_of_orders desc limit 10;
```

customer_city	no_of_orders
sao paulo	15540
rio de janeiro	6882
belo horizonte	2773
brasilia	2131
curitiba	1521
campinas	1444
porto alegre	1379
salvador	1245
guarulhos	1189
sao bernardo do campo	938

	customer_city	no_of_orders ▾
1.	sao paulo	15,540
2.	rio de janeiro	6,882
3.	belo horizonte	2,773
4.	brasilia	2,131
5.	curitiba	1,521
6.	campinas	1,444
7.	porto alegre	1,379
8.	salvador	1,245
9.	guarulhos	1,189
10.	sao bernardo do campo	938

1 - 10 / 10 < >



32. Top 10 states with the highest customers:

```
SELECT customer_state,COUNT(DISTINCT(customer_id)) AS NO_OF_CUSTOMERS FROM
Target.customers
GROUP BY customer_state ORDER BY NO_OF_CUSTOMERS DESC LIMIT 10;
```

customer_state	NO_OF_CUSTOMERS
SP	41746
RJ	12852
MG	11635
RS	5466
PR	5045
SC	3637
BA	3380
DF	2140
ES	2033
GO	2020

33. Customers with the highest number of orders:

```
select c.customer_unique_id, count(distinct(order_id)) as no_of_purchase from
Target.customers c, Target.orders o
where c.customer_id=o.customer_id
group by customer_unique_id order by no_of_purchase desc limit 10;
```

customer_unique_id	no_of_purchase
8d50f5eadf50201ccdcedfb9e2...	17
3e43e6105506432c953e165fb...	9
ca77025e7201e3b30c44b472f...	7
6469f99c1f9dfae7733b25662e...	7
1b6c7548a2a1f9037c1fd3ddfe...	7
47c1a3033b8b77b3ab6e109eb...	6
f0e310a6839dce9de1638e0fe...	6
12f5d6e1cbf93dafd9dcc19095...	6
de34b16117594161a6a89c50...	6
dc813062e0fc23409cd255f7f5...	6

34. Total numbers of orders, total numbers of orders delivered, and total numbers of orders canceled:

```
SELECT COUNT(DISTINCT(order_id)) as total_orders from Target.orders;
```

total_orders
99441

```sql
select count(distinct(order_id)) as orders_delivered from Target.orders where
order_status='delivered';
```

orders_delivered
96478

```sql
select count(distinct(order_id)) as orders_canceled from Target.orders where
order_status='canceled';
```

orders_canceled
625

35. How customers have rated the service?

```sql
select review_score, count(review_score) as cnt from Target.order_reviews group by
review_score ORDER BY review_score DESC;
```

review_score	cnt
5	57328
4	19142
3	8179
2	3151
1	11424

36. We can also check how customers' rating has improved/worsened over the entire period.

```sql
select '20'||year as year, review_score, count(review_score) as cnt from
(
select *, EXTRACT(day FROM review_creation_date) as year from Target.order_reviews
)
group by year, review_score
order by year, review_score desc;
```

year	review_score	cnt
2016	5	156
2016	4	51
2016	3	22
2016	2	8
2016	1	88
2017	5	24759
2017	4	8440
2017	3	3589
2017	2	1350
2017	1	4597
2018	5	32413
2018	4	10651
2018	3	4568
2018	2	1793
2018	1	6739

We can see that the rating has improved over the period of time. And, there is a significant increase in the review count in 2017 from 2016.

Analysis of Products

37. Total numbers of product categories available:

```
SELECT COUNT(DISTINCT(product_category)) as no_of_categories FROM Target.products;
```

no_of_categories
73

38. Top 10 categories with the most number of products:

```
SELECT product_category,COUNT(DISTINCT(product_id)) as no_of_products FROM
Target.products GROUP BY product_category ORDER BY no_of_products DESC LIMIT 10;
```

product_category	no_of_products
bed table bath	3029
sport leisure	2867
Furniture Decoration	2657
HEALTH BEAUTY	2444
housewares	2335
automotive	1900
computer accessories	1639
toys	1411
Watches present	1329
telephony	1134

39. Top 10 categories with the lowest number of products:

```
SELECT product_category,COUNT(DISTINCT(product_id)) as no_of_products FROM
Target.products GROUP BY product_category ORDER BY no_of_products asc LIMIT 10;
```

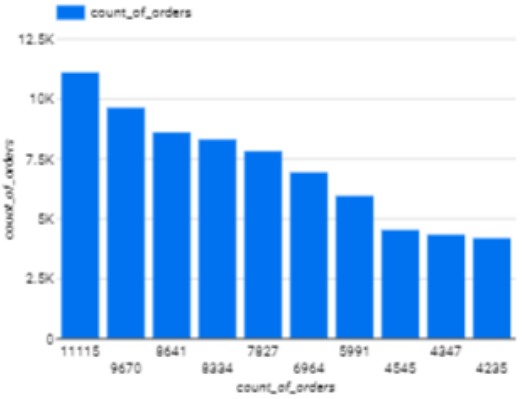
product_category	no_of_products
cds music dvds	1
insurance and services	2
PC Gamer	3
Fashion Children's Clothing	5
House Comfort 2	5
IMAGE IMPORT TABLETS	9
CITTE AND UPHACK FURNITURE	10
La Cuisine	10
Kitchen portable and food coach	10
Hygiene diapers	12

40. Top 10 highest-selling product categories in terms of no. of orders:

```
select product_category, count(o.order_id) as count_of_orders from Target.products
p, Target.orders o, Target.order_items i
where o.order_id=i.order_id
and i.product_id=p.product_id
group by product_category
order by count_of_orders desc limit 10;
```

	product_category	count_of_orders
1.	bed table bath	11,115
2.	HEALTH BEAUTY	9,670
3.	sport leisure	8,641
4.	Furniture Decoration	8,334
5.	computer accessories	7,827
6.	housewares	6,964
7.	Watches present	5,991
8.	telephony	4,545
9.	Garden tools	4,347
10.	automotive	4,235

1 - 10 / 10

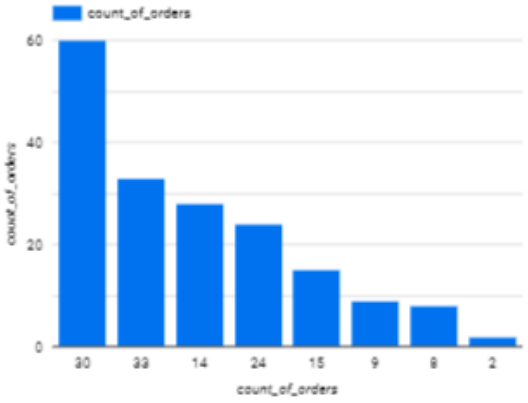


41. Bottom 10 selling product categories in terms no. of orders:

```
select product_category, count(o.order_id) as count_of_orders from Target.products
p, Target.orders o, Target.order_items i
where o.order_id=i.order_id
and i.product_id=p.product_id
group by product_category
order by count_of_orders asc limit 10;
```

	product_category	count_of_orders
1.	flowers	33
2.	Fashion Sport	30
3.	House Comfort 2	30
4.	Arts and Crafts	24
5.	Kitchen portable and food coach	15
6.	La Cuisine	14
7.	cds music dvds	14
8.	PC Gamer	9
9.	Fashion Children's Clothing	8
10.	insurance and services	2

1 - 10 / 10



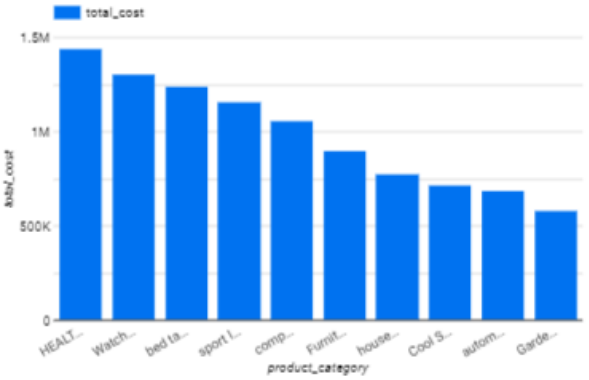
42. Top 10 highest selling products in terms of amount:

```
select product_category, round(sum(price+freight_value),2) as total_cost from
Target.products p, Target.orders o,
Target.order_items i
where o.order_id=i.order_id
```

```
and i.product_id=p.product_id
group by product_category
order by total_cost desc limit 10 ;
```

	product_category	total_cost
1.	HEALTH BEAUTY	1,441,248.07
2.	Watches present	1,305,541.61
3.	bed table bath	1,241,681.72
4.	sport leisure	1,156,656.48
5.	computer accessories	1,059,272.4
6.	Furniture Decoration	902,511.79
7.	housewares	778,397.77
8.	Cool Stuff	719,329.95
9.	automotive	685,384.32
10.	Garden tools	584,219.21

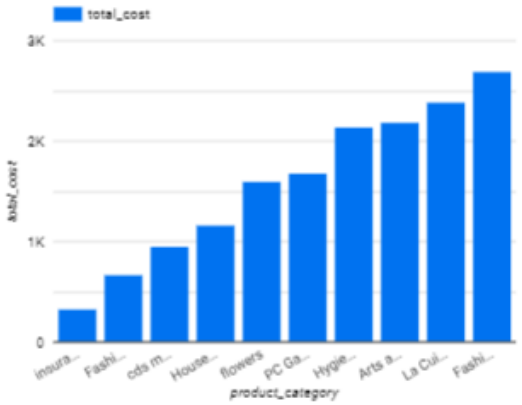
1 - 10 / 10



43. Bottom 10 selling products in terms of amount:

	product_category	total_cost
1.	insurance and services	324.51
2.	Fashion Children's Clothing	665.36
3.	cds music dvds	954.99
4.	House Comfort 2	1,170.58
5.	flowers	1,598.91
6.	PC Gamer	1,679.52
7.	Hygiene diapers	2,141.27
8.	Arts and Crafts	2,184.14
9.	La Cuisine	2,388.54
10.	Fashion Sport	2,697.64

1 - 10 / 10



From the above analysis done on product category and no. of orders and amount, we can see if there is a relation between no. of products in a category with their sales. We can observe the highest-selling product category also has the highest no. of products.

44. Reviews containing negative words like bad, dissatisfactory, worse, etc.:

```
SELECT count(review_id) as count from Target.order_reviews where
review_comment_title LIKE '%bad%' or review_comment_title LIKE '%Bad%'
or review_comment_title LIKE'%dissatisfac' or review_comment_title LIKE "%worse%";
```

count
104

45. Review comments containing positive words like good, great, recommend, satisfactory, etc.

```
SELECT count(review_comment_title) as count from Target.order_reviews where
review_comment_title LIKE "%ecommend%" or review_comment_title LIKE "%great%"
or review_comment_title LIKE "%good%" or review_comment_title LIKE "%satisfac%";
```

count
3079

Insights from the above analysis

- The data is for the period of 4th August 2016 to 17th October 2018. The sales are very less in 2016. It has picked only after 2017.
- The customer for Target who has purchased has been spread across 27 states and 4119 cities. We have a total of 8011 cities in Brazil from the geolocation table.
- The growth in sales has increased over the years. 2018 sales in terms of money are 139% higher than in 2017.
- We have the highest sales in the month of August, May, and July. For 2017, we have the highest sales in November and December.
- There is no particular date in the month that records the highest sales, but higher sales are observed in the second half of the month. In many cases after 20 the sales are more.
- Sao Paulo is the state with the highest number of sales.
- 22790, 24220, 22793, and 24230 are some of the area zip codes with the highest sales.
- 22790, 24220, 22793, and 24230 are the area with the highest number of customers buying from Target.
- RR, PB, RO, AC, and PI states with high avg freight value. SP, PR, MG, RJ, DF states with the lowest avg freight value. SP, RJ, MG, RS, and PR states with the highest total freight value.
- States with the lowest avg freight value has been among the states with highest total freight value i.e. low freight has resulted in more sales.
- RR, AP, AM, AL, and PA states with the highest avg estimated delivery time.
- AC, RO, AP, AM, and AR are the states where delivery has been fastest.

- AL, MA, SE, ES, and BA are the states where delivery has been slowest.
- Estimated delivery has improved over the entire duration. Although, it did not have shown impact on the review score.
- The review score has shown better results with delivery earlier than estimated.
- A total business of around 15 million has been done for the given time period.
- The highest order was done of 13.6 k.
- A total of 3095 sellers have registered themselves with Target. And, we have observed the states with more sellers have more sales.
- Most numbers of the sellers registered with Target are from Sao Paulo.
- Most of the customer purchases were from Sao Paulo and Rio de Janeiro. Sao Paulo is much ahead of other states.
- The trend in the buying of the product is more in the Afternoon or in the Night.
- The order count peaks during the mid-year period of May to August with the latter months having a count of orders above 10000.
- Most people do prefer to pay via credit cards(> 75%); UPI(20%) is the second most preferred choice of payment.
- From 2017 to 2018 a massive growth in money movement has been observed. The value of orders received in 2017 was around 3.6 million rose up to 8.6 million in 2018.
- A total of 99k orders were made, and 625 orders were canceled.
- More than 57% of the reviews have been given the top rating.
- The rating has improved with time and a significant increase has been observed in the review count over the period.
- There are 73 categories of products.
- Bed table bath, Sports leisure, furniture, and Health Beauty categories have the highest number of products. CDs/DVDs, insurance services, and PC gamers are among the ones with the lowest products available.
- Product categories with a higher number of products tend to have more sales.
- Health Beauty product category giving the highest sales in terms of money. And, Bed table bath is highest in terms of the number of orders.
- Insurance service, fashion children's clothing, and CDs, DVDs product category are the lowest in terms of the number of orders and amount.

Recommendations

- We have seen faster delivery time has shown an increase in the number of orders. So, Target can work on improving it further to achieve even better results.

- We have seen there is a better review from customers when the delivery is earlier than estimated. Target can work on it to speed up the delivery and a better customer review can help in the expansion of business.
- A relatively larger number of sales has been observed in August, May, and July. Also, in the year 2017, there is quite a number of sales in November and December. We can launch more products and give more offers in these months to further increase the number of sales.
- We have higher sales in the latter half of the month. We can make use of AI and try to offer personal recommendations basis their interest during the second half of the month.
- We also have seen top customers with the highest purchases; Target can put them into a special category and give them gifts and offers and try to build a personal relationship with them that can help them in retaining the top customers.
- Sao Paulo is the state with the highest number of sales. Target can invest something on studying the reason behind that and employ that technique in the other states. The seller quality, product quality, delivery time, and other factors which is making it happen in Sao Paulo can be employed in the other states to increase sales.
- The Zip code with the highest sales can be targeted further by advertising the brand in those areas by putting billboards, and hoardings.
- Target can reduce freight value to increase sales. As from the data we have seen the lowest avg freight value states have given more business.
- Most payments were done through credit cards. So, Target can tie up with more credit card partners to give more offers to the customers to increase sales. They can also set up their own payment network and credit card for a smoother and better experience for customers. UPI being the second most preferred payment method can also be targeted; Target can partner with more banks and payment networks to give customers a seamless experience to further increase their hold in the market.
- More sellers have resulted in more sales; Target can work to increase tie up with more sellers to expand their business.
- Product categories with more number of products have shown better results. Target can also see which sellers are giving more varieties for a product category and make partnerships with those sellers. Also, they can ask their sellers to work on increasing the number of products.
- Health Beauty has been the category with the highest sales in terms of amount; it can be further given importance to yield more profit.
- Looking at the product analysis, a few promising product category also can be focused on to increase their hold in the market.
- Bad reviews are very compared to good reviews, so Target can make sure to maintain its status quo to retain customers. The bad review can be studied and action can be taken to improve the customer experience.