BUSINESS CASE: TARGET
SALES ANALYSIS

Ques 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**

Table 1: Customers

Field name	Type	Mode
customer_id	STRING	NULLABLE
customer_unique_id	STRING	NULLABLE
customer_zip_code_prefix	INTEGER	NULLABLE
customer_city	STRING	NULLABLE
customer_state	STRING	NULLABLE

Table 2: Geolocation

Field name	Type	Mode
geolocation_zip_code_prefix	INTEGER	NULLABLE
geolocation_lat	FLOAT	NULLABLE
geolocation_lng	FLOAT	NULLABLE
geolocation_city	STRING	NULLABLE
geolocation_state	STRING	NULLABLE

Table 3: Order_Items

Field name	Type	Mode
order_id	STRING	NULLABLE
order_item_id	INTEGER	NULLABLE
product_id	STRING	NULLABLE
seller_id	STRING	NULLABLE
shipping_limit_date	TIMESTAMP	NULLABLE
price	FLOAT	NULLABLE
freight_value	FLOAT	NULLABLE

Table 4: Order_reviews

Field name	Type	Mode
review_id	STRING	NULLABLE
order_id	STRING	NULLABLE
review_score	INTEGER	NULLABLE
review_comment_title	STRING	NULLABLE
review_creation_date	TIMESTAMP	NULLABLE
review_answer_timestamp	TIMESTAMP	NULLABLE

Table 5: Orders

Field name	Type	Mode
order_id	STRING	NULLABLE
customer_id	STRING	NULLABLE
order_status	STRING	NULLABLE
order_purchase_timestamp	TIMESTAMP	NULLABLE
order_approved_at	TIMESTAMP	NULLABLE
order_delivered_carrier_date	TIMESTAMP	NULLABLE
order_delivered_customer_date	TIMESTAMP	NULLABLE
order_estimated_delivery_date	TIMESTAMP	NULLABLE

Table 6: Payments

Field name	Type	Mode
order_id	STRING	NULLABLE
payment_sequential	INTEGER	NULLABLE
payment_type	STRING	NULLABLE
payment_installments	INTEGER	NULLABLE
payment_value	FLOAT	NULLABLE

Table 7: products

Field name	Type	Mode
product_id	STRING	NULLABLE
product_category	STRING	NULLABLE
product_name_length	INTEGER	NULLABLE
product_description_length	INTEGER	NULLABLE
product_photos_qty	INTEGER	NULLABLE
product_weight_g	INTEGER	NULLABLE
product_length_cm	INTEGER	NULLABLE
product_height_cm	INTEGER	NULLABLE
product_width_cm	INTEGER	NULLABLE

Table 8: sellers

Field name	Type	Mode
seller_id	STRING	NULLABLE
seller_zip_code_prefix	INTEGER	NULLABLE
seller_city	STRING	NULLABLE
seller_state	STRING	NULLABLE

2) **Time period for which the data is given.**

```
select max(order_purchase_timestamp) as max_timestamp,
min(order_purchase_timestamp) as min_timestamp
from `TARGET.orders`;
```

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

As we can see above, the time period for which the data is given is coming out to be from 2016 to 2018.

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

```
select distinct c.customer_state as State, c.customer_city as City,
COUNT(o.order_id) as Total_orders
from `TARGET.customers` as c inner join `TARGET.orders` as o
on c.customer_id = o.customer_id
group by 1,2
order by 3 desc;
```

Row	State	City	Total_orders
1	SP	sao paulo	15540
2	RJ	rio de janeiro	6882
3	MG	belo horizonte	2773
4	DF	brasilia	2131
5	PR	curitiba	1521
6	SP	campinas	1444
7	RS	porto alegre	1379
8	BA	salvador	1245
9	SP	guarulhos	1189
10	SP	sao bernardo do campo	938

The above table shows the list of all the different states and cities of Brazil from where orders have been received by Target in the given time period.
As we can see, the maximum orders are received from capital of the most populous state, Sao Paulo.

Ques 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 peaks at specific months?

```
select *
from
 (select FORMAT_DATETIME('%b', DATETIME(ifnull(o.order_delivered_customer_date,o.order_estima
ted_delivery_date))) as purchase_month, count(o.order_id) as peak
from `TARGET.orders` as o inner join `TARGET.customers` as c
on o.customer_id = c.customer_id
where o.order_status = 'delivered'
group by 1)
order by peak desc
limit 10;
```

Row	purchase_month	peak
1	Aug	12616
2	May	10862
3	Jun	10054
4	Apr	9699
5	Jul	9299
6	Mar	9206
7	Dec	7210
8	Feb	7201
9	Jan	6880
10	Nov	4728

The peak performance is in the month of August, May and June while low performance is in the month of September, October and November.

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

```
select
(case
when extract(hour from order_purchase_timestamp) between 0 and 6
then 'DAWN'
```

```
when extract(hour from order_purchase_timestamp) between 7 and 11
then 'MORNING'
when extract(hour from order_purchase_timestamp) between 12 and 17
then 'EVENING'
else 'NIGHT'
end) as purchase_time , count(order_id) as Total_orders
from `TARGET.orders`
group by 1
order by 1;
```

Row	purchase_time	Total_orders
1	DAWN	5242
2	EVENING	38361
3	MORNING	21738
4	NIGHT	34100

The maximum purchases are made in the evening and at night.

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

1) Get month on month orders by states.

```
select c.customer_state, extract(year from o.order_purchase_timestamp) as Year, extract(month
 from o.order_purchase_timestamp) as Month, count(o.order_id) as total_orders
from `TARGET.orders` as o join `TARGET.customers` as c
on o.customer_id = c.customer_id
group by 2,3,1
order by 4 desc
limit 10;
```

Row	customer_state	Year	Month	total_orders
1	SP	2018	8	3253
2	SP	2018	5	3207
3	SP	2018	4	3059
4	SP	2018	1	3052
5	SP	2018	3	3037
6	SP	2017	11	3012
7	SP	2018	7	2777
8	SP	2018	6	2773
9	SP	2018	2	2703
10	SP	2017	12	2357

The maximum orders that Target received is in the month of August of 2018 and the most orders are made by customers of SP state.

2) Distribution of customers across the states in Brazil.

```sql
select customer_state, count(customer_id) as total_customers, round(count(di
stinct customer_id) * 100/ (select count(*) from `TARGET.customers`),2) as D
istribution_per_state
from `TARGET.customers`
group by 1
order by 2 desc
limit 10;
```

Row	customer_state	total_customers	Distribution_per_state
1	SP	41746	41.98
2	RJ	12852	12.92
3	MG	11635	11.7
4	RS	5466	5.5
5	PR	5045	5.07
6	SC	3637	3.66
7	BA	3380	3.4
8	DF	2140	2.15
9	ES	2033	2.04
10	GO	2020	2.03

The highest distribution of customers is in Sao Paulo state with almost 42% of total customers who actively make purchases from Target stores.
The lowest distribution of customers is in Roraima with only 0.05 % of total customers.

Ques 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.**

```sql
select round(((total_2018 - total_2017) * 100 / total_2017),2) as percentage
_inc_2017_2018
from
(select sum(case when years = 2017 then total_payment_value
else 0
end) as total_2017 , sum(case when years = 2018 then total_payment_value
else 0
end) as total_2018
from
(select extract(month from o.order_purchase_timestamp) as months , extract(y
ear from o.order_purchase_timestamp) as years, sum (p.payment_value) as tota
l_payment_value
from `TARGET.payments` as p join `TARGET.orders` as o
on p.order_id = o.order_id
where extract(month from order_purchase_timestamp) IN (1,2,3,4,5,6,7,8) and
extract(year from order_purchase_timestamp) in (2017, 2018)
group by years, months) as A) as B
```

Row	percentage_inc_2017_2018
1	136.98

The increase in the cost of orders from 2017 to 2018 is approximately equal to 137 %.

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

```
select c.customer_state, round(avg(i.price),2) as avg_price , round(sum(i.pr
ice),2) as sum_price, round(avg(i.freight_value),2) as avg_freight_value, ro
und(sum(i.freight_value),2) as sum_freight_value
from `TARGET.order_items` as i join `TARGET.orders` as o
on i.order_id = o.order_id
join `TARGET.customers` as c
on o.customer_id = c.customer_id
group by c.customer_state
Order by 1
limit 10;
```

Row	customer_state	avg_price	sum_price	avg_freight_value	sum_freight_value
1	AC	173.73	15982.95	40.07	3686.75
2	AL	180.89	80314.81	35.84	15914.59
3	AM	135.5	22356.84	33.21	5478.89
4	AP	164.32	13474.3	34.01	2788.5
5	BA	134.6	511349.99	26.36	100156.68
6	CE	153.76	227254.71	32.71	48351.59
7	DF	125.77	302603.94	21.04	50625.5
8	ES	121.91	275037.31	22.06	49764.6
9	GO	126.27	294591.95	22.77	53114.98
10	MA	145.2	119648.22	38.26	31523.77

Taget can look for ways to reduce the freight value in order to attract more customers.

Ques 5. Analysis on sales, freight and delivery time

1) Calculate days between purchasing, delivering and estimated delivery

```
select order_id, date_diff(order_delivered_customer_date,order_purchase_time
stamp,day ) as days_to_delivering , date_diff(order_estimated_delivery_date,
order_purchase_timestamp, day) as days_to_estimated_delivery
from `TARGET.orders`
where order_delivered_customer_date is not null
order by 1
limit 10;
```

Row	order_id	days_to_delivering	days_to_estimated_delivery
1	00010242fe8c5a6d1ba2dd792...	7	15
2	00018f77f2f0320c557190d7a1...	16	18
3	000229ec398224ef6ca0657da...	7	21
4	00024acbcdf0a6daa1e931b03...	6	11
5	00042b26cf59d7ce69dfabb4e...	25	40
6	00048cc3ae777c65dbb7d2a06...	6	21
7	00054e8431b9d7675808bcb8...	8	24
8	000576fe39319847cbb9d288c...	5	20
9	0005a1a1728c9d785b8e2b08...	9	9
10	0005f50442cb953dcd1d21e1f...	2	20

If we order the above query by column: days_to_delivering in desc order, we can see that there are a good number of orders that took a lot of days to get delivered. The below screenshot shows the difference in the days when it was estimated to get delivered vs the days it actually took to deliver the order.

Row	order_id	days_to_delivering	days_to_estimated_delivery
1	ca07593549f1816d26a572e06...	209	28
2	1b3190b2dfa9d789e1f14c05b...	208	19
3	440d0d17af552815d15a9e41a...	195	30
4	0f4519c5f1c541ddec9f21b3bd...	194	32
5	285ab9426d6982034523a855f...	194	28
6	2fb597c2f772eca01b1f5c561b...	194	39
7	47b40429ed8cce3aee9199792...	191	15
8	2fe324febf907e3ea3f2aa9650...	189	22
9	2d7561026d542c8dbd8f0daea...	188	28
10	437222e3fd1b07396f1d9ba8c...	187	42

2) Find time_to_delivery & diff_estimated_delivery.

```
select customer_id, order_id,  date_diff(order_delivered_customer_date, orde
r_purchase_timestamp, day) as time_to_delivery, date_diff(order_estimated_de
livery_date, order_delivered_customer_date, day) as diff_estimated_delivery
from `TARGET.orders`
order by 1,2
limit 10;
```

Row	customer_id	order_id	time_to_delivery	diff_estimated_delivery
1	00012a2ce6f8dcda20d059ce9...	5f79b5b0931d63f1a42989eb6...	13	5
2	000161a058600d5901f007fab...	a44895d095d7e0702b6a162fa...	9	9
3	0001fd6190edaaf884bcaf3d49...	316a104623542e4d75189bb3...	5	15
4	0002414f95344307404f0ace7...	5825ce2e88d5346438686b0b...	28	0
5	000379cdec625522490c315e7...	0ab7fb08086d4af9141453c91...	11	4
6	0004164d20a9e969af783496f...	cd3558a10d854487b4f907e9b...	8	13
7	000419c5494106c306a97b56...	07f6c3baf9ac86865b60f640c4...	45	-26
8	00046a560d407e99b969756e...	8c3d752c5c02227878fae49ae...	8	16
9	00050bf6e01e69d5c0fd612f1b...	fa906f338cee30a984d0945b3...	15	10
10	000598caf2ef4117407665ac3...	9b961b894e797f63622137ff7e...	9	16

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

```sql
select c.customer_state, round(avg(i.freight_value),2) as avg_freight_value,
  round(avg(date_diff(o.order_delivered_customer_date, o.order_purchase_time
stamp, day)),2) as avg_time_to_delivery, round(avg(date_diff(o.order_estimat
ed_delivery_date, o.order_delivered_customer_date, day)),2) as avg_diff_esti
mated_delivery
from `TARGET.customers` as c join `TARGET.orders` as o
on c.customer_id = o.customer_id
join `TARGET.order_items` as i
on o.order_id = i.order_id
group by 1
order by 1
limit 10;
```

Row	customer_state	avg_freight_value	avg_time_to_delivery	avg_diff_estimated_delivery
1	AC	40.07	20.33	20.01
2	AL	35.84	23.99	7.98
3	AM	33.21	25.96	18.98
4	AP	34.01	27.75	17.44
5	BA	26.36	18.77	10.12
6	CE	32.71	20.54	10.26
7	DF	21.04	12.5	11.27
8	ES	22.06	15.19	9.77

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

For getting highest value:

```sql
select c.customer_state, round(avg(i.freight_value),2) as avg_freight_value
from `TARGET.order_items` as i join `TARGET.orders` as o
on i.order_id = o.order_id
join `TARGET.customers` as c
on o.customer_id = c.customer_id
group by 1
order by 2 desc
limit 5;
```

Row	customer_state	avg_freight_value
1	RR	42.98
2	PB	42.72
3	RO	41.07
4	AC	40.07
5	PI	39.15

For getting lowest value:

```sql
select c.customer_state, round(avg(i.freight_value),2) as avg_freight_value
```

```
from `TARGET.order_items` as i join `TARGET.orders` as o
on i.order_id = o.order_id
join `TARGET.customers` as c
on o.customer_id = c.customer_id
group by 1
order by 2
limit 5
```

Row	customer_state	avg_freight_value
1	SP	15.15
2	PR	20.53
3	MG	20.63
4	RJ	20.96
5	DF	21.04

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

For getting highest value:

```
select c.customer_state,  round(avg(date_diff(o.order_delivered_customer_date,
o.order_purchase_timestamp, day)),2) as avg_time_to_delivery
from `TARGET.orders` as o join `TARGET.customers` as c
on o.customer_id = c.customer_id
group by 1
order by 2 desc
limit 5;
```

Row	customer_state	avg_time_to_delivery
1	RR	28.98
2	AP	26.73
3	AM	25.99
4	AL	24.04
5	PA	23.32

For getting lowest value:

```
select c.customer_state,  round(avg(date_diff(o.order_delivered_customer_date,
o.order_purchase_timestamp, day)),2) as avg_time_to_delivery
from `TARGET.orders` as o join `TARGET.customers` as c
on o.customer_id = c.customer_id
group by 1
order by 2
limit 5;
```

Row	customer_state	avg_time_to_delivery
1	SP	8.3
2	PR	11.53
3	MG	11.54
4	DF	12.51
5	SC	14.48

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

For fast delivery:

```
select c.customer_state, round(avg(t.diff_in_days),2) as average_diff_in_day
s
from `TARGET.customers` as c
join
(select o.*, timestamp_diff(o.order_estimated_delivery_date, o.order_deliver
ed_customer_date,day) as diff_in_days,
from `TARGET.orders` as o
where o.order_status = 'delivered') as t
on c.customer_id = t.customer_id
group by 1
order by 2 desc
limit 5
```

Row	customer_state	average_diff_in_days
1	AC	19.76
2	RO	19.13
3	AP	18.73
4	AM	18.61
5	RR	16.41

For not so fast delivery:

```
select c.customer_state, round(avg(t.diff_in_days),2) as average_diff_in_days
from `TARGET.customers` as c
join
(select o.*, timestamp_diff(o.order_estimated_delivery_date, o.order_delivered_
customer_date,day) as diff_in_days,
from `TARGET.orders` as o
where o.order_status = 'delivered') as t
on c.customer_id = t.customer_id
group by 1
order by 2
limit 5
```

Row	customer_state	average_diff_in_days
1	AL	7.95
2	MA	8.77
3	SE	9.17
4	ES	9.62
5	BA	9.93

Ques 6. Payment type analysis:

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

```
select count(t.orders) as Total_orders, t.Month, t.payment_type
from
(select o.order_id as orders, extract(month from o.order_purchase_timestamp) as Month, p.paym
ent_type as payment_type
from `TARGET.payments` as p join `TARGET.orders` as o
on p.order_id = o.order_id) as t
group by Month, payment_type
order by 3,2
limit 10;
```

Row	Total_orders	Month	payment_type
1	1715	1	UPI
2	1723	2	UPI
3	1942	3	UPI
4	1783	4	UPI
5	2035	5	UPI
6	1807	6	UPI
7	2074	7	UPI
8	2077	8	UPI
9	903	9	UPI
10	1056	10	UPI

If we further calculate, we find that:
Payment of only 1529 orders is made by debit cards and 19784 transactions are done by UPI. The maximum transactions are done by credit cards (total orders = 76795)

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

```
select payment_installments, count (distinct order_id) as count_orders
from `TARGET.payments`
group by 1
order by 2 desc
limit 10;
```

Row	payment_installments	count_orders
1	1	49060
2	2	12389
3	3	10443
4	4	7088
5	10	5315
6	5	5234
7	8	4253
8	6	3916
9	7	1623
10	9	644

Maximum customers have made payment in the 1st installment itself.

Insights and Recommendations:

1) The highest distribution of customers is in Sao Paulo state with almost 42% of total customers who actively make purchases from Target stores.
2) The lowest distribution of customers is in Roraima with only 0.05 % of total customers.
3) The most active customers are in SP state, specially in the month of May, June, and August. Therefore, Target can come up with more attractive offers to further increase the sales in these months.
4) During September, October and November, the sales are not up to the mark. Hence, some form of marketing should be done to direct the customers into buying in these months as well.
5) Target should also work on lowering the freight value of the orders in order to attract more customers.
6) The speed of order delivery in AL, MA, SE, ES, BA is very slow. If that gets resolved, more customers will engage in purchasing from Target.
7) Payment of only 1529 orders is made by debit cards and 19784 transactions are done by UPI. The maximum transactions are done by credit cards (total orders paid by credit cards = 76795)