checking the structure & characteristics of the dataset:

QUESTIONS

1. **Data type of all columns in the "customers" table**.

ANS:

query:

```
select
     column_name,data_type
from
    `digital-layout-
408815.target_data.INFORMATION_SCHEMA.COLUMNS`
where table_name='customers'
```

snapshot:

JOB INFORMATION | RESULTS | CHART | PREVIEW

w	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

INSIGHTS: There are two types of data types have been used one is “STRING” and another one is “INTEGER”

RECOMMENDATIONS: N/A

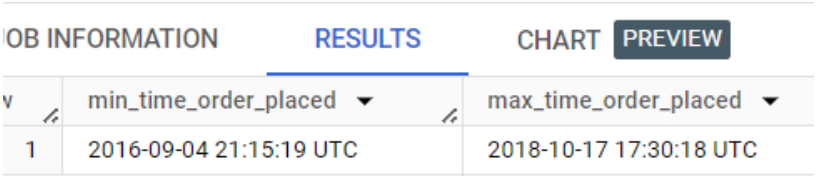
2.Get the time range between which the orders were placed.

ANS:

query:

```
select
      min(order_purchase_timestamp) as `earliest_order_placed`,
      max(order_purchase_timestamp) as `latest_order_placed`
from
   `target_data.orders`
```

Snapshot:



INSIGHTS: Target earliest order and latest order was placed in these two timestamps these two orders came in consecutive months but different years there many be any seasonal sale going on by observing these dates.

RECOMMENDATIONS: N/A

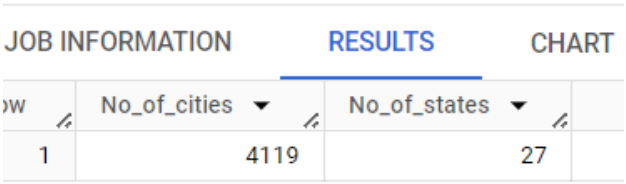
3.Count the Cities & States of customers who ordered during the given period

ANS:

query:

```
select
      count( distinct customer_city) as `No_of_cities`,
      count(distinct customer_state) as `No_of_states`
from `target_data.customers`
```

Snapshot:



INSIGHTS: There are 27 states and 4119 cities by observing above data orders are placed by different cities and states it concludes that this company reached to some extent but far more to go

RECOMMENDATIONS: This company needs to focus on other areas and need to do more advertisements and adds to attract customers by using market strategy

2.In-depth Exploration:

4.Is there a growing trend in the no. of orders placed over the past years?

ANS

Query:

```
select
  extract(year from order_purchase_timestamp) as `order_year`,
  count(order_id) as `no_of_orders`
from `target_data.orders`
group by order_year
order by order_year asc
```

snapshot:

JOB INFORMATION | RESULTS | CHART

Row	order_year	no_of_orders
1	2016	329
2	2017	45101
3	2018	54011

INSIGHTS: From above results we can observe that there is a growing trend in the number of orders placed over past years because in 2016 there is only 326 orders a sudden rise in number of orders in 2017 which went 45101 orders and 2018 got 54011 a tremendous rise in orders within two years.

RECOMMENDATIONS: In my point of view the company is doing great work by improving year by year because results says it all a tremendous growth in business to continue this company need to increase its employees and staff to manage this high traffic

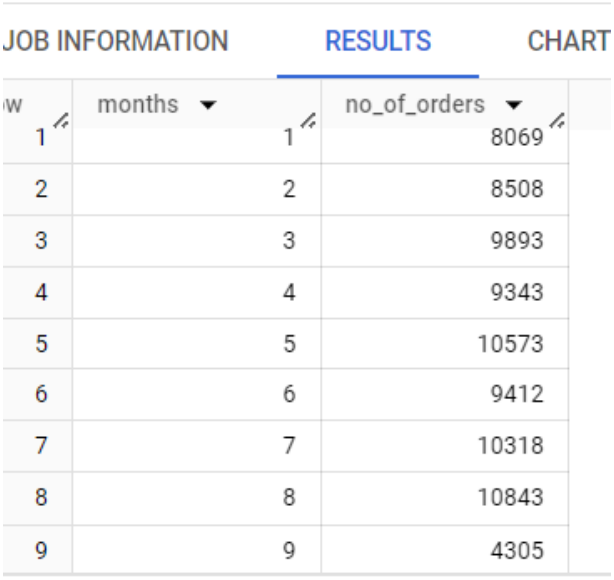
5.Can we see some kind of monthly seasonality in terms of the no. of orders being placed?

ANS

Query:

```sql
select extract(month from order_purchase_timestamp) as months,
       count(order_id) as no_of_orders
from `target_data.orders`
group by months
order by months asc
```

snapshot:



JOB INFORMATION | RESULTS | CHART

w	months	no_of_orders
1	1	8069
2	2	8508
3	3	9893
4	4	9343
5	5	10573
6	6	9412
7	7	10318
8	8	10843
9	9	4305

INSIGHTS: From above results we can observe that when we compare orders in months consecutive increase of sales from first month to eight months but sudden drop of sales from ninth month to the rest of the year. Customers mostly ordered only between 1 to 8 months less orders in remaining months. This is due to any reasons like New year, festivals discounts summer season affect where some people don't want to expose their faces to sunlight, so they prefer online orders which safe and secure to get desire products from offline store which the target company is providing. Assuming that Most of cold drinks and fruits got high demand in summer but low sales in winter. This is what we call seasonality orders .

RECOMMENDATIONS: By keeping this result in view company need to focus on seasonality products in stock which will paly vital role in increasing sales. Company also need to take advantage of this situation to increase the non-seasonality products by giving combo pack with seasonality products by providing some discounts on combo packs for example assume that one wants to buy cold drink in summer he searched for it but seeing the combo pack of cold drink along with biscuits and chips he will get attracted to it and suddenly mindset of customer change and he will buy total combo pack of it. This is what company needs to do to increase their sales

6.During what time of the day, do the Brazilian customers mostly place their orders? (Dawn, Morning, Afternoon or Night)

- **0-6 hrs : Dawn**
- **7-12 hrs : Mornings**
- **13-18 hrs : Afternoon**
- **19-23 hrs : Night**

ANS:

Query:

```
select count(order_id) as no_of_orders,
       (case
           when t.hours between 0 and 6 then "Dawn"
           when t.hours between 7 and 12 then "Morning"
           when t.hours between 13 and 18 then "Afternoon"
           else "Night" end) as period_of_day
from
    (select *,extract(hour from order_purchase_timestamp) as
hours

    from `target_data.orders`
    ) as t
group by period_of_day
order by no_of_orders asc
```

snapshot:

JOB INFORMATION | RESULTS | CHART

Row	no_of_orders	period_of_day
1	5242	Dawn
2	27733	Morning
3	28331	Night
4	38135	Afternoon

INSIGHTS: By observing above results Brazilian customers mostly placed their orders in Afternoon followed by night and morning where there is a rapid decrease in Dawn because most of the people were in sleep and inactive or doing some other daily activities. This result shows that most of the customers are interested in shopping at afternoon period comparing to other periods there was very less customers in Dawn period . Night and Morning are in second and third places

RECOMMENDATIONS:

Target company needs to make availability of more staff in company at afternoon to manage the heavy traffic. This makes easy for customers to save their time also needs to place less staff in dawn period due to less orders. Company also needs to take care about the stock in afternoon period

otherwise it will lead to lack of stock and customers gets disappointed with this service directly leads to decrease in number of customers.

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

7.Get the month on month no. of orders placed in each state.

ANS

QUERY:

```
select customer_state,extract(month from
order_purchase_timestamp) as months,
  count(order_id) as no_of_orders
from
    `target_data.orders` as o
inner join `target_data.customers` as c on
o.customer_id=c.customer_id
group by months,customer_state
order by no_of_orders desc
```

snapshot:

Row	customer_state	months	no_of_orders
1	SP	8	4982
2	SP	5	4632
3	SP	7	4381
4	SP	6	4104
5	SP	3	4047
6	SP	4	3967
7	SP	2	3357
8	SP	1	3351
9	SP	11	3012

INSIGHTS: From above SP customer_state has highest number of orders placed in all months followed by RJ state and MG state with second and third highest RR state has lowest order rate when comparing to other states

RECOMMENDATIONS: The above result is based on several factors due to distance between company and states,low population in some states, unnoticed in some states .Company needs to extend as much as make advertisements to promote their sales

8.How are the customers distributed across all the states?

ANS

Query:

```
select
      customer_state,
      count(customer_id) as no_of_customers
from `target_data.customers`
group by customer_state
order by no_of_customers desc
```

snapshot:

Row	customer_state	no_of_customers
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
11	PE	1652
12	CE	1336

INSIGHTS: From we can observe the nearly half of the customers are from SP state and followed by RJ state where RR state and AP state has least number of customers. We can see a huge difference between first highest number of customers and second highest number of customers . We can consider that SP state is the most important state for this company.

Recommendations: Target needs to focus on least number of customers state like RR,AP,AC and AM states this is due to many reasons target needs to conduct some kind of survey in this areas to find out the reasons for this solve the problems which are

mentioned in the survey also have to make more and more advertisements on their services. This is due to distance also target need to expand their branches as much as to attract the customers which can make easy for some states

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

9.Get the % increase in the cost of orders from year 2017 to 2018 (include months between Jan to Aug only).
You can use the "payment_value" column in the payments table to get the cost of orders.

ANS:

Query:

```
select round((((t.total_cost_2018-
t.total_cost_2017)/total_cost_2017)*100),2) as percentage_increse
from
    (select
       sum(
        case
        when extract(year from order_purchase_timestamp)=2017
        and
        extract(month from order_purchase_timestamp)  between 1
and 8
        then payment_value else 0 end
        ) as total_cost_2017,
      sum(
        case
        when extract(year from order_purchase_timestamp)=2018
        and
        extract(month from order_purchase_timestamp) between 1
and 8
        then payment_value else 0 end
        ) as total_cost_2018
   from
    target_data.orders as o inner join target_data.payments as p
on o.order_id=p.order_id) as t
```

snapshot:

Row	percentage_increse
1	136.98

INSIGHTS: By observing above results there is 136.98% increase in the cost of orders from year 2017 to 2018 including months between Jan to Aug .This is a tremendous profit in 2018 target is doing good by creating a massive profit in just one year of span this results excluding last four months the sales of last four months are not as much good to consider but these eight months made a big difference in profit percentage.

RECOMMENDATIONS: Target company needs do more and more profits by extending their services in world wide. It has to develop day by day because trend is changing day by day they need to come with different types of ideas to tackle the difficulty situations in future they need to improve their marketing by using new technologies like AI, chat gpt this will help the company to stand out in the market.

10.Calculate the Total & Average value of order price for each state

ANS:

Query:

```
select customer_state,round(avg(payment_value),2) as avg_value,
                      round(sum(payment_value),2) as total_value
from `target_data.payments` as p
          inner join `target_data.orders` as o on
p.order_id=o.order_id
          inner join `target_data.customers` as c on
c.customer_id=o.customer_id
group by customer_state
order by total_value desc
```

snapshot:

Row	customer_state	avg_value	total_value
1	SP	137.5	5998226.96
2	RJ	158.53	2144379.69
3	MG	154.71	1872257.26
4	RS	157.18	890898.54
5	PR	154.15	811156.38
6	SC	165.98	623086.43
7	BA	170.82	616645.82
8	DF	161.13	355141.08
9	GO	165.76	350092.31
10	ES	154.71	325967.55
11	PE	187.99	324850.44
12	CE	199.9	279464.03

INSIGHTS: From above results SP customer state has highest total value and less average value followed by RJ state there is huge difference between first and second total value. This shows the SP state contributing in large value where RR state has lowest value in terms of total value here total value refers sum of the prices of all orders by no of orders

RECOMMENDATIONS: Company needs to focus on low level states to generate a new customers from these states. It needs to implement some marketing strategies by providing discounts on occasions and giving some gifts in by taking a lucky draw or lottery which help to attract the customers

11.Calculate the Total & Average value of order freight for each state.

ANS:

Query:

```
select customer_state,round(avg(freight_value),2) as
avg_freight_value,

round(sum(freight_value),2) as total_freight_value
from `target_data.order_items` as oi
inner join `target_data.orders` as o on oi.order_id=o.order_id
inner join `target_data.customers` as c on
c.customer_id=o.customer_id
group by customer_state
```

```
order by total_freight_value desc
```

snapshot:

Row	customer_state	avg_freight_value	total_freight_value
1	SP	15.15	718723.07
2	RJ	20.96	305589.31
3	MG	20.63	270853.46
4	RS	21.74	135522.74
5	PR	20.53	117851.68
6	BA	26.36	100156.68
7	SC	21.47	89660.26
8	PE	32.92	59449.66
9	GO	22.77	53114.98
10	DF	21.04	50625.5
11	ES	22.06	49764.6
12	CE	32.71	48351.59

Insights: From above results we can conclude that average freight value is less for SP state followed by RJ state and total freight value is greater for these two states where RR and AP states has highest average freight value and lowest total freight value, freight value means price for transporting item from one place to other it is like delivery charges it states that SP state is near from this company which will cost less freight price where RR state is very far from this company it leads to increase in cost of freight price increase in freight price is directly proportional to decrease in number of orders

RECOMMENDATIONS: Company needs to extend its services by expanding their branches in some far away states which leads to increase customers in these states and it will also reduce the long delivering period to short period it will also decrease in average freight values and increase in total freight values

5.Analysis based on sales, freight and delivery time

12.Find the no. of days taken to deliver each order from the order's purchase date as delivery time.
Also, calculate the difference (in days) between the estimated & actual delivery date of an order.

ANS:

Query:

```
Select
order_id,datetime_diff(order_delivered_customer_date,order_purcha
se_timestamp,day) as `no_of_days_taken_to_delivery`,

       datetime_diff(order_delivered_customer_date,order_estimate
d_delivery_date,day) as `diff_bwt_delivered_and_estimated_days`
from `target_data.orders`
```

snapshot:

Row	order_id	no_of_days_taken_to_delivery	diff_bwt_delivered_and_estimated_days
1	1950d777989f6a877539f5379...	30	12
2	2c45c33d2f9cb8ff8b1c86cc28...	30	-28
3	65d1e226dfaeb8cdc42f66542...	35	-16
4	635c894d068ac37e6e03dc54e...	30	-1
5	3b97562c3aee8bdedcb5c2e45...	32	0
6	68f47f50f04c4cb6774570cfde...	29	-1
7	276e9ec344d3bf029ff83a161c...	43	4
8	54e1a3c2b97fb0809da548a59...	40	4
9	fd04fa4105ee8045f6a0139ca5...	37	1
10	302bb8109d097a9fc6e9cefc5...	33	5

INSIGHTS: By observing the results above table gives the information second column indicates the number of days taken to delivery the order and third column indicates that difference between order delivered date and order estimated date. We can see both positive and negative values in third column, positive values states that order is delivered after the estimated date and negative values state that order is delivered prior to the estimated value

RECOMMENDATIONS:

For some orders time taken to delivery the order is very long which cause the customers to lose interest on buying other products and one more thing is need to consider average difference between delivered date and estimated date is 10 to 20 days this shows that company is not providing accurate estimated date to the orders all the above cases are depend on some factors like products are not in stock and due to lack of delivery agents company needs to hire more and more employees to delivery the product in time

13.Find out the top 5 states with the highest & lowest average freight value

ANS:

Query:

```
select customer_state,round(avg(freight_value),2) as
avg_freight_value

from `target_data.order_items` as oi
inner join `target_data.orders` as o on oi.order_id=o.order_id
inner join `target_data.customers` as c on
c.customer_id=o.customer_id
group by customer_state
order by avg_freight_value asc
```

snapshot: Top 5 lowest freight values

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	SC	21.47
7	RS	21.74
8	ES	22.06

For Top 5 highest freight values are

```
select customer_state,round(avg(freight_value),2) as
avg_freight_value

from `target_data.order_items` as oi
inner join `target_data.orders` as o on oi.order_id=o.order_id
inner join `target_data.customers` as c on
c.customer_id=o.customer_id
group by customer_state
order by avg_freight_value desc
```

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
6	MA	38.26
7	TO	37.25
8	SE	36.65

INSIGHTS: The above results give top five average freight value and top lowest average freight values. Freight value can be defined has price charged for transporting product from one place to other place by observing above results RR state has highest average value and followed by PB,RO and AC states where SP state has lowest average freight values

RECOMMENDATIONS: By observing above data the states which has highest average freight value are states are long distance from the orders placed which will the major cause for increasing freight values this indirectly effecting the customers where lowest freight values states are short distance from order place this will not effect the customers. Company should focus on this issue because there are very less number of customers from highest average freight values, so company should look alternative to over come this by providing services from short distances by expanding their branches as much as.

14.Find out the top 5 states with the highest & lowest average delivery time.

ANS:

Query:

```sql
select customer_state,round(avg(no_of_days_taken_to_delivery),2)
as avg_delivered_days
from

   (select order_id,customer_state,
            datetime_diff(order_delivered_customer_date,order_pu
rchase_timestamp,day) as no_of_days_taken_to_delivery

       from `target_data.orders`  as o
       inner join `target_data.customers` as c on
o.customer_id=c.customer_id) as t
 group by customer_state
 order by avg_delivered_days desc
```

snapshot: Top 5 states with highest average delivery time

Row	customer_state	avg_delivered_days
1	RR	28.98
2	AP	26.73
3	AM	25.99
4	AL	24.04
5	PA	23.32
6	MA	21.12
7	SE	21.03
8	CE	20.82
9	AC	20.64
10	PB	19.95
11	PI	18.99
12	RO	18.91

For Top 5 states with lowest average delivery time

```sql
select customer_state,round(avg(no_of_days_taken_to_delivery),2)
as avg_delivered_days
from

      (select order_id,customer_state,
            datetime_diff(order_delivered_customer_date,order_pu
rchase_timestamp,day) as no_of_days_taken_to_delivery

       from `target_data.orders`  as o
       inner join `target_data.customers` as c on
o.customer_id=c.customer_id) as t
 group by customer_state
 order by avg_delivered_days asc
```

Row	customer_state	avg_delivered_days
1	SP	8.3
2	PR	11.53
3	MG	11.54
4	DF	12.51
5	SC	14.48
6	RS	14.82
7	RJ	14.85
8	GO	15.15
9	MS	15.19
10	ES	15.33
11	TO	17.23
12	MT	17.59

INSIGHTS: The above results shows that top state with lowest time taken to delivery the order examples are SP,PR and MG states these states are taking less time to receive the product where some states like RR,AP and AM states are taking long time to receive their orders

RECOMMENDATIONS: By observing above data this data indirectly effecting the number of customers because of the long delivery time customers are looking into alternative options this leads to decrease in customer numbers target should look into this issue and need to work on providing some fast delivery services to far away areas which is the one option to increase the customers from these states.

15.Find out the top 5 states where the order delivery is really fast as compared to the estimated date of delivery.
You can use the difference between the averages of actual & estimated delivery date to figure out how fast the delivery was for each state

ANS:

QUERY:

```
select
customer_state,round(avg(t.diff_bwt_delivered_and_estimated_days)
,2) as avg_fast_delivered_days

from

      (select order_id,customer_state,
            datetime_diff(order_delivered_customer_date,order_pu
rchase_timestamp,day) as no_of_days_taken_to_delivery,
           datetime_diff(order_delivered_customer_date,order_est
imated_delivery_date,day) as
           diff_bwt_delivered_and_estimated_days
       from `target_data.orders`  as o
       inner join `target_data.customers` as c on
o.customer_id=c.customer_id) as t

 group by customer_state
 order by avg_fast_delivered_days
```

snapshot:

Row	customer_state	avg_fast_delivered_days
1	AC	-19.76
2	RO	-19.13
3	AP	-18.73
4	AM	-18.61
5	RR	-16.41
6	MT	-13.43
7	PA	-13.19
8	RS	-12.98
9	RN	-12.76
10	PE	-12.4
11	PB	-12.37
12	PR	-12.36

INSIGHTS: From above results we can conclude that AC state is the top fast delivered days with 19 days prior to the estimated days negative symbol refers that order is delivered prior to the estimated date and positive symbol refers order is delivered after the estimated date after AC state followed by RO,AP,AM and RR states

RECOMMENDATIONS: When we take the average values of difference between delivered date and estimated date. Target giving its best by delivering its orders prior to the estimated dates, but when we compare individual orders some orders are getting long delay in delivery time the estimated dates which they are giving are very long from order purchase dates target need to look into this the top states are in the above result that does not mean they are receiving their order fast it is fast when compared to the estimated dates because some estimated dates very long company should look into this issue.

6.Analysis based on the payments

16.Find the month on month no. of orders placed using different payment types.

ANS:

Query:

```sql
select
    extract(month from order_purchase_timestamp) as months,
    payment_type,
    count(o.order_id) as no_of_orders

from `target_data.orders` as o
inner join `target_data.payments` as p o.order_id=p.order_id
group by months,payment_type
order by no_of_orders desc
```

snapshot:

Row	months	payment_type	no_of_orders
1	5	credit_card	8350
2	8	credit_card	8269
3	7	credit_card	7841
4	3	credit_card	7707
5	4	credit_card	7301
6	6	credit_card	7276
7	2	credit_card	6609
8	1	credit_card	6103
9	11	credit_card	5897
10	12	credit_card	4378
11	10	credit_card	3778
12	9	credit_card	3286

INSIGHTS: By observing above results on month on month most of the orders are placed through credit cards there are total five payment types are used like credit card, UPI, debit card, vouchers but most of the people used credit card for their purchases after credit card some people made payment using UPI mode very least people used debit card and vouchers

RECOMMENDATIONS: The company need focus on the data because they need to make arrangements for the customers according to their convenient payment methods it makes easy and time saving process for customers, target should provide discounts on particular credit cards by collaborating with that banks which will generate the additional revenue or the Target need provide their personal cards which is help for doing payments for only regular customers it helps to avoid some taxes

17.Find the no. of orders placed on the basis of the payment installments that have been paid.

ANS:

Query:

```
select payment_installments,
      count(order_id) as no_of_orders
from `target_data.payments`
group by payment_installments
order by no_of_orders desc
```

snapshot:

Row	payment_installment	no_of_orders
1	1	52546
2	2	12413
3	3	10461
4	4	7098
5	10	5328
6	5	5239
7	8	4268
8	6	3920
9	7	1626
10	9	644
11	12	133
12	15	74

INSIGHTS: By observing the above results we can see that most of the people order products using only single payment only few of them went to two to twenty four months of installments process because to avoid extra or additional charges people preferring single payment method.

RECOMMENDATIONS: By analysing the results people are preferring only one time payment due to discounts or any other reasons, in some cases discounts are applicable

only on one time payment process. Target should also focus on giving discounts on installments which will attract more and more customers because most of the people are employees they will also prefer installments when they want buy some big products.