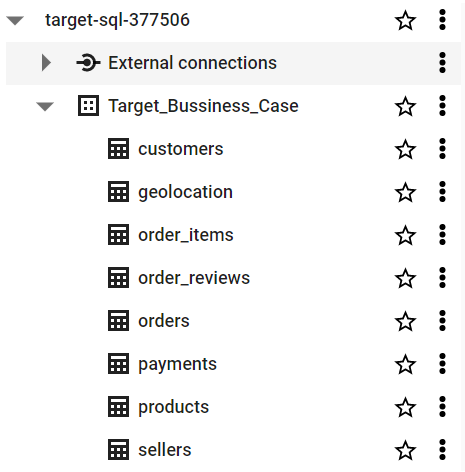
Target Business Case Study

- Target is one of the world's most recognized brands and one of America's leading retailers.
- This business case has information of 100,000 orders from 2016-2018 made at Target in Brazil

1. Import the dataset and do usual exploratory analysis steps like checking the structure & characteristics of the dataset

- Datasets given for performing the analysis:

1. customers.csv
2. geolocation.csv
3. order_items.csv
4. payments.csv
5. reviews.csv
6. orders.csv
7. products.csv
8. sellers.csv



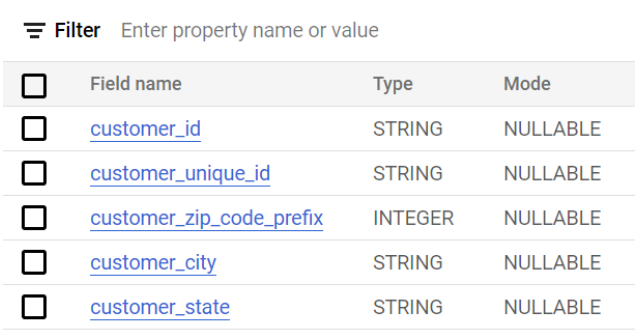
Dataset imported in BigQuerry

2. Data type of columns in a table

- Each feature of different csv's is described below:

➢ customers.csv

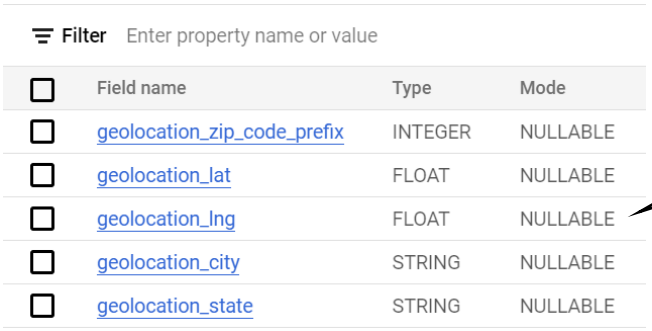
Features	Description	Datatype
customer_id	Id of the consumer who made the purchase.	STRING
customer_unique_id	Unique Id of the consumer.	STRING
customer_zip_code_prefix	Zip Code of the location of the consumer.	INTEGER
customer_city	Name of the City from where order is made.	STRING
customer_state	State Code from where order is made (Ex- Sao Paulo-SP).	STRING



SCHEMA for customers.csv from BigQuerry

➢ geolocation.csv

Features	Description	Datatype
geolocation_zip_code_prefix	first 5 digits of zip code	INTEGER
geolocation_lat	latitude	FLOAT
geolocation_lng	longitude	FLOAT
geolocation_city	city name	STRING
geolocation_state	state	STRING

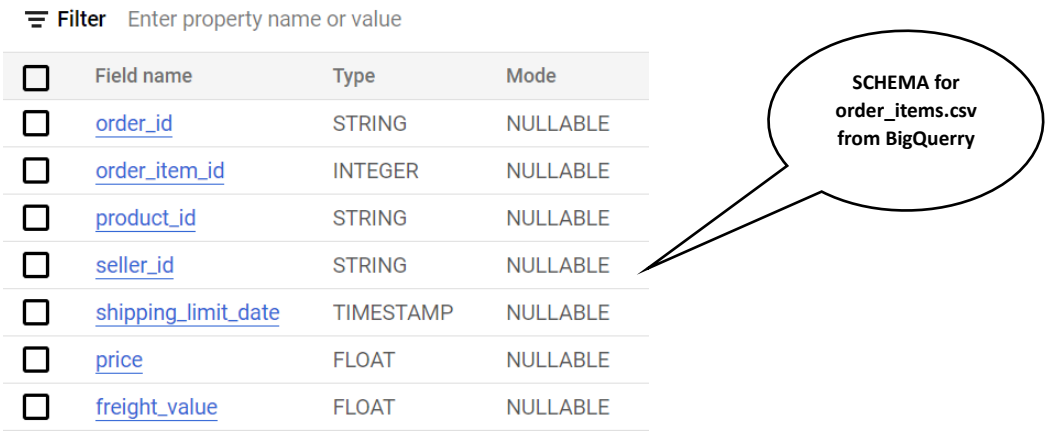


Filter Enter property name or value

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

➢ order_items.csv

Features	Description	Datatype
order_id	A unique id of order made by the consumers.	STRING
order_item_id	A Unique id given to each item ordered in the order.	INTEGER
product_id	A unique id given to each product available on the site.	STRING
seller_id	Unique Id of the seller registered in Target.	STRING
shipping_limit_date	The date before which shipping of the ordered product must be completed.	TIMESTAMP
price	Actual price of the products ordered.	FLOAT
freight_value	Price rate at which a product is delivered from one point to another.	FLOAT



Filter Enter property name or value

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

➢ payments.csv

Features	Description	Datatype
order_id	A unique id of order made by the consumers.	STRING
payment_sequential	sequences of the payments made in case of EMI.	INTEGER
payment_type	mode of payment used. (Ex-Credit Card)	STRING
payment_installments	number of installments in case of EMI purchase.	INTEGER
payment_value	Total amount paid for the purchase order.	FLOAT

Filter Enter property name or value

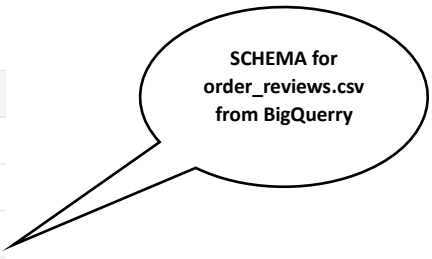
Field name	Type	Mode
order_id	STRING	NULLABLE
payment_sequential	INTEGER	NULLABLE
payment_type	STRING	NULLABLE
payment_installments	INTEGER	NULLABLE
payment_value	FLOAT	NULLABLE

➢ order_reviews.csv

Features	Description	Datatype
review_id	Id of the review given on the product ordered by the order id.	STRING
order_id	A unique id of order made by the consumers.	STRING
review_score	review score given by the customer for each order on the scale of 1–5.	INTEGER
review_comment_title	Title of the review	STRING
review_creation_date	Timestamp of the review when it is created.	TIMESTAMP
review_answer_timestamp	Timestamp of the review answered.	TIMESTAMP

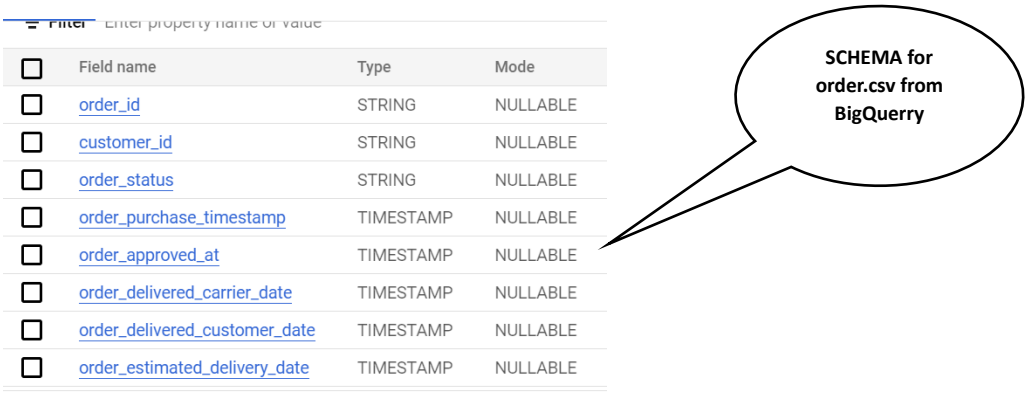
Filter Enter property name or value

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



➢ orders.csv

Features	Description	Datatype
order_id	A unique id of order made by the consumers.	STRING
customer_id	Id of the consumer who made the purchase.	STRING
order_status	status of the order made i.e delivered, shipped etc.	STRING
order_purchase_timestamp	Timestamp of the purchase.	TIMESTAMP
order_delivered_carrier_date	delivery date at which carrier made the delivery.	TIMESTAMP
order_delivered_customer_date	date at which customer got the product.	TIMESTAMP
order_estimated_delivery_date	estimated delivery date of the products.	TIMESTAMP



➢ products.csv

Features	Description	Datatype
product_id	A unique identifier for the proposed project.	STRING
product_category_name	Name of the product category	STRING
product_name_lenght	length of the string which specifies the name given to the products ordered.	INTEGER

product_description_lenght	length of the description written for each product ordered on the site.	INTEGER
product_photos_qty	Number of photos of each product ordered available on the shopping portal.	INTEGER
product_weight_g	Weight of the products ordered in grams.	INTEGER
product_length_cm	Length of the products ordered in centimeters.	INTEGER
product_height_cm	Height of the products ordered in centimeters.	INTEGER
product_width_cm	width of the product ordered in centimeters.	INTEGER

Filter Enter property name or value

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

- **sellers.csv**

Features	Description	Datatype
seller_id	Unique Id of the seller registered	STRING
seller_zip_code_prefix	Zip Code of the location of the seller.	INTEGER
seller_city	Name of the City of the seller.	STRING
seller_state	State Code (Ex- sao paulo-SP)	STRING

Filter Enter property name or value

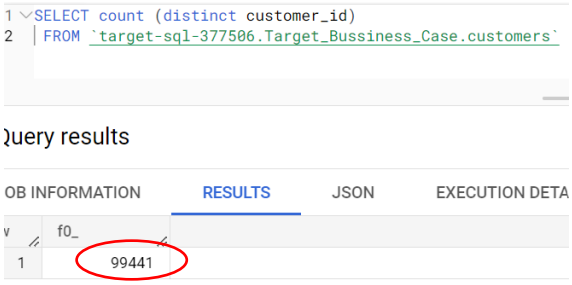
Field name	Type	Mode
seller_id	STRING	NULLABLE
seller_zip_code_prefix	INTEGER	NULLABLE
seller_city	STRING	NULLABLE
seller_state	STRING	NULLABLE

SCHEMA for selers.csv from BigQuerry

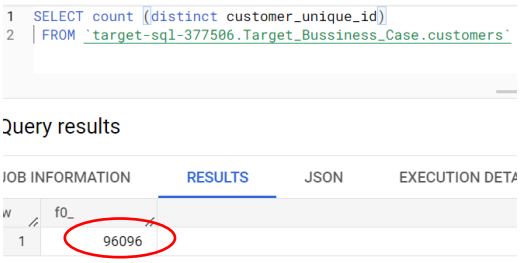
Cities and States of customers ordered during the given period

➢ Exploratory Analysis of the data

- Here data is available for a total of 99441 customers.



- There are 96096 unique customer id's.



- There are customers from over **27 states, 4119 cities and 14994 unique zip codes** of Brazil
- There is data for **99441 orders**, of which
  - **Sao Paulo** has the **most orders** i.e., **41746.**
  - **Roraima has the least orders i.e., 46**

- Out of these 99441 orders, **1107 orders have been shipped, 625 orders cancelled and 96478 has been delivered**

```
SELECT
order_status,
COUNT(1) AS Number_of_Orders
FROM `target-sql-377506.Target_Bussiness_Case.orders`
GROUP BY order_status
```

Query results

JOB INFORMATION | RESULTS | JSON | EXECUTION

	order_status	Number_of_Orders
1	created	5
2	shipped	1107
3	approved	2
4	canceled	625
5	invoiced	314
6	delivered	96478
7	processing	301

3. Time period for which the data is given

- We have a data for a time period of **25 months**.

```
SELECT
DATE_DIFF(MAX(DATE(order_delivered_customer_date)),MIN(DATE(order_purchase_timestamp)),month) AS
time_period_in_months
FROM `target-sql-377506.Target_Bussiness_Case.orders`
```

Press Alt+F1

Query results SAVE RESULTS EXPLO

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS | EXECUTION GRAPH PREVIEW

	time_period_in_months
1	25

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?

- There is a year-on-year increase in the revenue and number of orders which indicates that the there is an emerging trend in e-commerce in Brazil.
- To support this

```sql
SELECT
EXTRACT(YEAR FROM order_purchase_timestamp) as _year,
ROUND(SUM(price), 2) AS revenue,
COUNT(o.order_id) AS number_of_Orders
FROM
`target-sql-377506.Target_Bussiness_Case.orders` as o
JOIN
`target-sql-377506.Target_Bussiness_Case.order_items` AS ois
ON o.order_id = ois.order_id
GROUP BY _year
ORDER BY _year;
```

Query results

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS

Row	_year	revenue	number_of_Orders
1	2016	49785.92	370
2	2017	6155806.98	50864
3	2018	7386050.8	61416

- From this we can conclude that there is **19% increase** in revenue from **2017 to 2018**.

This table shows the average orders per month.

Average number of orders are higher during **November** month, **September and October month average orders are comparatively low,** in **May and July and august have higher average orders** compare to other months.

Row	month	avgerage_orders
1	1	4034.5
2	2	4254.0
3	3	4946.5
4	4	4671.5
5	5	5286.5
6	6	4706.0
7	7	5159.0
8	8	5421.5
9	9	1435.0
10	10	1653.0
11	11	7544.0
12	12	2837.0

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

- From this we can see that the maximum orders are placed in the **Afternoon** and least in the **Dawn**

Row	time	total_order
1	Afternoon	38361
2	Evening	24161
3	Morning	20507
4	night	14491
5	Dawn	1921

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

1. **Get month on month orders by states**
2. **Distribution of customers across the states in Brazil**

Number of orders per state

Row	customer_state	number_of_orde	state	region
1	SP	41746	São Paulo	Southeast
2	RJ	12852	Rio de Janeiro	Southeast
3	MG	11635	Minas Gerais	Southeast
4	RS	5466	Rio Grande do Sul	South
5	PR	5045	Paraná	South
6	SC	3637	Santa Catarina	South
7	BA	3380	Bahia	Northeast
8	DF	2140	Distrito Federal	Center West
9	ES	2033	Espírito Santo	Southeast
10	GO	2020	Goiás	Center West
11	PE	1652	Pernambuco	Northeast
12	CE	1336	Ceará	Northeast
13	PA	975	Pará	North

Revenue and Average Sales per stat

Row	customer_state	revenue_per_sta	avg_sale_per_st	state	region
1	SP	5998226.95...	137.504629...	São Paulo	Southeast
2	RJ	2144379.68...	158.525888...	Rio de Janeiro	Southeast
3	MG	1872257.26...	154.706433...	Minas Gerais	Southeast
4	RS	890898.539...	157.180405...	Rio Grande do Sul	South
5	PR	811156.379...	154.153625...	Paraná	South
6	SC	623086.429...	165.979336...	Santa Catarina	South
7	BA	616645.820...	170.816016...	Bahia	Northeast
8	DF	355141.080...	161.134791...	Distrito Federal	Center West
9	GO	350092.310...	165.763404...	Goiás	Center West
10	ES	325967.55	154.706953...	Espírito Santo	Southeast
11	PE	324850.440...	187.992152...	Pernambuco	Northeast
12	CE	279464.029...	199.902739...	Ceará	Northeast
13	PA	218295.849...	215.920722...	Pará	North

Number of customer and revenue per state

Row	customer_city	number_of_cust	revenue
1	sao paulo	15540	2203373.09...
2	rio de janeiro	6882	1161927.35...
3	belo horizonte	2773	421765.120...
4	brasilia	2131	354216.780...
5	curitiba	1521	247392.480...
6	campinas	1444	216248.429...
7	porto alegre	1379	224731.419...
8	salvador	1245	218071.499...
9	guarulhos	1189	165121.99
10	sao bernardo do campo	938	120434.839...
11	niteroi	849	139996.990...
12	santo andre	797	105627.089...
13	osasco	746	94358.7199...

4. Impact on Economy: Analyse 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
  avg(payment_value) as Average_Revenue_2017
from
  `target-sql-377506.Target_Bussiness_Case.payments` as p
  JOIN `target-sql-377506.Target_Bussiness_Case.orders` as o
  on p.order_id=o.order_id
where
  extract(year FROM date(order_purchase_timestamp))=2017
  and extract(month from (order_purchase_timestamp)) between 01 and 08;
```

Query results

	Average_Revenue_2017
1	150.42524373744448

```sql
select
  avg(payment_value) as Average_Revenue_2018
from
  `target-sql-377506.Target_Bussiness_Case.payments` as p
  JOIN `target-sql-377506.Target_Bussiness_Case.orders` as o
  on p.order_id=o.order_id
where
  extract(year FROM date(order_purchase_timestamp))=2018
  and extract(month from (order_purchase_timestamp)) between 01 and 08;
```

Query results

	Average_Revenue_2018
1	155.27696830073944

Average Revenue for 2017 from Jan to August is 150.43

Average Revenue for 2018 from Jan to August is 155.28

So, % increase can be interpreted as 3.22%

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

```sql
select
  cust.state,cust.region,
  round(avg(payment_value),2) as Avg_Price,
  round(sum(payment_value),2) as Total_Price,
  round(avg(freight_value),2) as Avg_Freigt_Charge,
  round(sum(freight_value),2)  as Total_Freight_Charge
from
  `target-sql-377506.Target_Bussiness_Case.payments` as pay
  join `target-sql-377506.Target_Bussiness_Case.order_items` as itm
  on pay.order_id=itm.order_id
  join `target-sql-377506.Target_Bussiness_Case.orders` as ord
  on ord.order_id=itm.order_id
  join `target-sql-377506.Target_Bussiness_Case.customers` as cust
  on ord.customer_id=cust.customer_id
group by cust.customer_id,cust.state,cust.region
order by 2 desc
```

ow	state	region	Avg_Price	Total_Price	Avg_Freigt_Char	Total_Freight_Ch
1	São Paulo	Southeast	153.27	7597209.66	15.2	753351.18
2	Rio de Janeiro	Southeast	180.68	2769347.44	21.1	323413.95
3	Minas Gerais	Southeast	170.56	2326151.64	20.63	281301.31
4	Espírito Santo	Southeast	173.57	405805.34	21.98	51392.57
5	Paraná	South	178.56	1064603.99	20.58	122669.69
6	Santa Catarina	South	182.79	786343.71	21.44	92216.36
7	Rio Grande do Sul	South	176.89	1147277.0	21.83	141579.69
8	Bahia	Northeast	196.99	797410.36	26.32	106538.62
9	Maranhão	Northeast	235.27	108566.27	38.26	33290.33

Sao Paulo has the lowest average price.

5. Analysis on sales, freight and delivery time

1. Calculate days between purchasing, delivering and estimated delivery

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

- time_to_delivery = order_purchase_timestamp-order_delivered_customer_date
- diff_estimated_delivery = order_estimated_delivery_date-order_delivered_customer_date

```sql
SELECT
    order_id,
    date_diff(order_delivered_customer_date,order_purchase_timestamp,day) as time_to_delivery,
    date_diff(order_estimated_delivery_date,order_delivered_customer_date,day)  as diff_estimated_delivery
FROM
    `target-sql-377506.Target_Bussiness_Case.orders`
```

Query results

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS | EXECUTION GRAPH PREVIEW

w	order_id	time_to_delivery	diff_estimated_c
1	1950d777989f6a877539f5379…	30	-12
2	2c45c33d2f9cb8ff8b1c86cc28…	30	28
3	65d1e226dfaeb8cdc42f66542…	35	16
4	635c894d068ac37e6e03dc54e…	30	1
5	3b97562c3aee8bdedcb5c2e45…	32	0
6	68f47f50f04c4cb6774570cfde…	29	1

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

```sql
SELECT
    c.state,c.customer_state,
    round(avg(date_diff(order_delivered_customer_date,order_purchase_timestamp,day)),2) as time_to_delivery,
    round(avg(date_diff(order_estimated_delivery_date,order_delivered_customer_date,day)),2) as diff_estimated_delivery,
    round(avg(oi.freight_value),2) avg_fright_value
FROM
`target-sql-377506.Target_Bussiness_Case.orders` as o
JOIN
`target-sql-377506.Target_Bussiness_Case.customers` as c
on o.customer_id = c.customer_id
JOIN
`target-sql-377506.Target_Bussiness_Case.order_items` as oi
on oi.order_id = o.order_id
group by
c.state,c.customer_state
```

Row	state	customer_state	time_to_delivery	diff_estimated_c	avg_freight_valu
1	MatoGrosso	MT	17.51	13.64	28.17
2	Maranhão	MA	21.2	9.11	38.26
3	Alagoas	AL	23.99	7.98	35.84
4	São Paulo	SP	8.26	10.27	15.15
5	Minas Gerais	MG	11.52	12.4	20.63
6	Pernambuco	PE	17.79	12.55	32.92
7	Rio de Janeiro	RJ	14.69	11.14	20.96

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

- Top 5 states with Lowest Freight Value

Row	state	customer_state	time_to_delivery	diff_estimated_c	avg_freight_valu
1	São Paulo	SP	8.26	10.27	15.15
2	Paraná	PR	11.48	12.53	20.53
3	Minas Gerais	MG	11.52	12.4	20.63
4	Rio de Janeiro	RJ	14.69	11.14	20.96
5	Distrito Federal	DF	12.5	11.27	21.04

- Top 5 states with Highest Freight Value

Row	state	customer_state	time_to_delivery	diff_estimated_c	avg_freight_valu
1	Roraima	RR	27.83	17.43	42.98
2	Paraíba	PB	20.12	12.15	42.72
3	Rondônia	RO	19.28	19.08	41.07
4	Acre	AC	20.33	20.01	40.07
5	Piauí	PI	18.93	10.68	39.15

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

- Top 5 states with Lowest Average Time to Delivery

Row	state	customer_state	time_to_delivery	diff_estimated_c	avg_freight_valu
1	São Paulo	SP	8.26	10.27	15.15
2	Paraná	PR	11.48	12.53	20.53
3	Minas Gerais	MG	11.52	12.4	20.63
4	Distrito Federal	DF	12.5	11.27	21.04
5	Santa Catarina	SC	14.52	10.67	21.47

- Top 5 states with Highest Average Time to Delivery

Row	state	customer_state	time_to_delivery	diff_estimated_c	avg_freight_valu
1	Roraima	RR	27.83	17.43	42.98
2	Amapá	AP	27.75	17.44	34.01
3	Amazonas	AM	25.96	18.98	33.21
4	Alagoas	AL	23.99	7.98	35.84
5	Pará	PA	23.3	13.37	35.83

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

- Top 5 states with Really fast compared to estimated delivery date

Row	state	customer_state	time_to_delivery	diff_estimated_c	avg_freight_valu
1	São Paulo	SP	8.26	10.27	15.15
2	Paraná	PR	11.48	12.53	20.53
3	Minas Gerais	MG	11.52	12.4	20.63
4	Rio de Janeiro	RJ	14.69	11.14	20.96
5	Distrito Federal	DF	12.5	11.27	21.04

- Top 5 states with not so fast compared to estimated delivery date

Row	state	customer_state	time_to_delivery	diff_estimated_c	avg_freight_valu
1	Roraima	RR	27.83	17.43	42.98
2	Paraíba	PB	20.12	12.15	42.72
3	Rondônia	RO	19.28	19.08	41.07
4	Acre	AC	20.33	20.01	40.07
5	Piauí	PI	18.93	10.68	39.15

6. Payment type analysis:

1. Month over Month count of orders for different payment types
2. Count of orders based on the no. of payment instalments

```
select
    payment_type,
    count(order_id) as order_cnt
from
    `target-sql-377506.Target_Bussiness_Case.payments`
group by
    payment_type
order by
    order_cnt desc
```

Row	payment_type	order_cnt
1	credit_card	76795
2	UPI	19784
3	voucher	5775
4	debit_card	1529
5	not_defined	3

```
select
  p.payment_type,
  extract(month from date(o.order_purchase_timestamp)) as month,
  count(o.order_id)
from `target-sql-377506.Target_Bussiness_Case.payments` p join `target-sql-377506.Target_Bussiness_Case.orders` o
on o.order_id=p.order_id
group by p.payment_type,
         extract(month from date(o.order_purchase_timestamp))
```

Query results

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS | EXECUTION GRAPH PREVIEW

ow	payment_type	month	f0_
1	credit_card	5	8350
2	credit_card	4	7301
3	voucher	1	477
4	voucher	4	572
5	voucher	10	318

```sql
select P.payment_installments,
       count(o.order_id) as order_cnt
from `target-sql-377506.Target_Bussiness_Case.payments` as P
JOIN `target-sql-377506.Target_Bussiness_Case.orders` as o
on o.order_id=P.order_id
group by P.payment_installments
order by 1 desc
```

Query results

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS

ow	payment_installr	order_cnt
1	24	18
2	23	1
3	22	1
4	21	3
5	20	17
6	18	27

Actionable Insights:

- Insights and Recommendations:
- We have 99,441 customers of data available.
- We have 96096 number of Unique Customers ids.
- 14994 different locations of customers
- Customers are from different 4119 cities and 27 states from Brazil.
- total 99441 customers are there in given data.
- from total 99441 orders, 1107 are shipped ,625 were cancelled, 96478 are delivered.
- Total 3095 sellers who are from 611 different cities and 23 states in Brazil and from 2246 different areas as per zip-code data.
- São Paulo state has the highest numbers of sellers in country.
- Analysis of sales and revenue as per time:
- Time period for which the data is given is 25 months.
- compare to 2017, revenue has increased in 2018 by 21%.
- Average number of orders are higher during November month, September and October month average orders are comparatively low, in May and July and august have higher average orders compare to other months.
- Tuesday, Monday and Wednesdays have relatively higher number of orders.

- Increasing trend:
- There is an increasing trend in orders, trend sustains during 2018. There a slight fall we can observe during October 2017 following with a great hike in November month and again a fall in end of December 2017 and January 2018.
- we can observe the trend of increasing orders with time and also for revenue.
- we can observe there's 815% growth increased in terms of orders and 707% growth increment in terms of revenue in January from 2017 to 2018.
- growth rate for July and august in 2017 to 2018 is relatively very low!
- 2017-february, 2017-march,2017-november were the highest growing sale month compare to its previous month.
- States São Paulo, Paraná, Minas Gerais, Distrito Federal, Santa Catarina and Rio de Janeiro are some of the states having faster delivery time relatively.
- Alagoas, Amazonas, Amama, Pará and Roraima are some states have very slow delivery time relatively.
- Region and State vice Analysis:
- São Paulo, Rio de Janeiro, Minas Gerais, Rio Grande do Sul and Paraná are top 5 highest orders states and also generating highest revenue. more than 80% of orders are coming from south, southeast and northeast Brazil. 90% of the revenue is coming from south, southeast and northeast Brazil.

Recommendations

Recommendations :

- From the distribution and statistical analysis, we can observe the average time to complete the delivery is 12 days. which should be reduced to at least half, as due to high competition in e-commerce market, it is vital to do so.
- If we look at Top states where delivery is really slow compared to estimated date, they are all from north Brazil region. Delivering faster in the north states may create and increase new customers and revenue from north.

- It was observed an increasing trend in revenue and orders over time, yet during October and January sales are decreasing probably after Festival Sales. Introducing possible discount on not so running product can help sell more products during those low going months.

- It is observed that the states where delivery is slow has a low order rate so there is need to strengthen the logistics to boost the revenue from those states.

- Afternoon is the time when there is a peak in the orders so we need to keep a check at staffing in the afternoon and also run special promotions for other time of the day.