Brazilian E-Commerce Data Analysis

This data contains 100k orders information between 2016-2018. Below analysis made by orders, geolocation, payments, seller, and products datasets. The aim of this project is finding the Top 10 products and sellers, analysis of orders by their geolocation and obtain information about Brazilian's online e-commerce profiles.

Data source: Kaggle

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**

CUSTOMERS

```
SELECT table_name,
       column_name,
       data_type
FROM   Ecommerce.INFORMATION_SCHEMA.COLUMNS
WHERE  table_schema = 'Ecommerce'
       AND table_name = 'customers';
```

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS

Row	table_name	column_name	data_type
1	customers	customer_id	STRING
2	customers	customer_unique_id	STRING
3	customers	customer_zip_code_prefix	INT64
4	customers	customer_city	STRING
5	customers	customer_state	STRING

GEOLOCATION

```
SELECT table_name,
       column_name,
       data_type
FROM   Ecommerce.INFORMATION_SCHEMA.COLUMNS
WHERE  table_schema = 'Ecommerce'
       AND table_name = 'geolocation';
```

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS

Row	table_name	column_name	data_type
1	geolocation	geolocation_zip_code_prefix	INT64
2	geolocation	geolocation_lat	FLOAT64
3	geolocation	geolocation_lng	FLOAT64
4	geolocation	geolocation_city	STRING
5	geolocation	geolocation_state	STRING

ORDER_ITEMS

```
SELECT table_name,
       column_name,
       data_type
FROM   Ecommerce.INFORMATION_SCHEMA.COLUMNS
WHERE  table_schema = 'Ecommerce'
       AND table_name = 'order_items';
```

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS

Row	table_name	column_name	data_type
1	order_items	order_id	STRING
2	order_items	order_item_id	INT64
3	order_items	product_id	STRING
4	order_items	seller_id	STRING
5	order_items	shipping_limit_date	TIMESTAMP
6	order_items	price	FLOAT64
7	order_items	freight_value	FLOAT64

ORDER_REVIEWS

```
SELECT table_name,
       column_name,
       data_type
FROM   Ecommerce.INFORMATION_SCHEMA.COLUMNS
WHERE  table_schema = 'Ecommerce'
       AND table_name = 'order_reviews';
```

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS

Row	table_name	column_name	data_type
1	order_reviews	review_id	STRING
2	order_reviews	order_id	STRING
3	order_reviews	review_score	INT64
4	order_reviews	review_comment_title	STRING
5	order_reviews	review_creation_date	TIMESTAMP
6	order_reviews	review_answer_timestamp	TIMESTAMP

ORDERS

```
SELECT table_name,
       column_name,
       data_type
FROM   Ecommerce.INFORMATION_SCHEMA.COLUMNS
WHERE  table_schema = 'Ecommerce'
       AND table_name = 'orders';
```

Query results

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS

Row	table_name	column_name	data_type
1	orders	order_id	STRING
2	orders	customer_id	STRING
3	orders	order_status	STRING
4	orders	order_purchase_timestamp	TIMESTAMP
5	orders	order_approved_at	TIMESTAMP
6	orders	order_delivered_carrier_date	TIMESTAMP
7	orders	order_delivered_customer_date	TIMESTAMP
8	orders	order_estimated_delivery_date	TIMESTAMP

PAYMENTS

```
SELECT table_name,
       column_name,
       data_type
FROM   Ecommerce.INFORMATION_SCHEMA.COLUMNS
WHERE  table_schema = 'Ecommerce'
       AND table_name = 'payments';
```

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS

Row	table_name	column_name	data_type
1	payments	order_id	STRING
2	payments	payment_sequential	INT64
3	payments	payment_type	STRING
4	payments	payment_installments	INT64
5	payments	payment_value	FLOAT64

PRODUCTS

```
SELECT table_name,
       column_name,
       data_type
FROM   Ecommerce.INFORMATION_SCHEMA.COLUMNS
WHERE  table_schema = 'Ecommerce'
       AND table_name = 'products';
```

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS

Row	table_name	column_name	data_type
1	products	product_id	STRING
2	products	product_category	STRING
3	products	product_name_length	INT64
4	products	product_description_length	INT64
5	products	product_photos_qty	INT64
6	products	product_weight_g	INT64
7	products	product_length_cm	INT64
8	products	product_height_cm	INT64
9	products	product_width_cm	INT64

SELLERS

```
SELECT table_name,
       column_name,
       data_type
FROM   Ecommerce.INFORMATION_SCHEMA.COLUMNS
WHERE  table_schema = 'Ecommerce'
       AND table_name = 'sellers';
```

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS

Row	table_name	column_name	data_type
1	sellers	seller_id	STRING
2	sellers	seller_zip_code_prefix	INT64
3	sellers	seller_city	STRING
4	sellers	seller_state	STRING

2. Time period for which the data is given

```
SELECT Min(order_purchase_timestamp)       AS start_date,
       Max(order_estimated_delivery_date) AS end_date
FROM   `Ecommerce.orders`;
```

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS

Row	start_date	end_date
1	2016-09-04 21:15:19 UTC	2018-11-12 00:00:00 UTC

Time Period for which data given is from 2016 Sept to 2018 Nov.

3. Cities and States covered in the dataset

CITY

```
SELECT DISTINCT geolocation_city AS City
FROM   `Ecommerce.geolocation`
```

Row	City
1	aracaju
2	riachuelo
3	nossa senhora do socorro
4	barra dos coqueiros
5	itaporanga d'ajuda
6	sao cristovao
7	são cristóvão
8	santo amaro das brotas
9	pirambu
10	laranjeiras

```
SELECT Count(DISTINCT geolocation_city) AS TotalCities
FROM   `Ecommerce.geolocation`
```

JOB INFORMATION

Row	TotalCities
1	8011

STATE

```
SELECT DISTINCT geolocation_state AS State
FROM   `Ecommerce.geolocation`
```

Row	State
1	SE
2	AL
3	PI
4	AP
5	AM
6	RR
7	AC
8	RO
9	TO
10	BA

```
SELECT Count(DISTINCT geolocation_state) AS TotalStates
FROM   `Ecommerce.geolocation`
```

JOB INFORMATION

Row	TotalStates
1	27

There are total 8011 unique cities and 27 states covered in the dataset.

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?**

SALES ACCORDING TO YEAR

```
SELECT Year,
       count_of_orders,
       Round(( count_of_orders * 100 ) / ( Sum(count_of_orders)
                                             OVER () ), 2) AS
       percentage_of_orders
FROM   (SELECT Count(order_id)                                AS count_of
_orders,
               Extract (Year FROM order_purchase_timestamp) AS Year,
       FROM   Ecommerce.orders
       WHERE  order_status NOT IN ( 'canceled', 'unavailable' )
       GROUP  BY Year)x
ORDER  BY count_of_orders DESC;
```

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS

Row	Year	count_of_orders	percentage_of_orders
1	2018	53532	54.51
2	2017	44379	45.19
3	2016	296	0.3

From the above result we can see that sales increased every year from 2016 to 2018, but it significantly increased from 2016 to 2017. Around 54 % of purchase made in year 2018.

SALES ACCORDING TO YEAR & MONTHS

```
SELECT month_name,
       year,
       count_of_orders,
       Round(( count_of_orders * 100 ) / ( Sum(count_of_orders)
                                             OVER () ), 2) AS
       percentage_of_orders
FROM   (SELECT Count(order_id)                                AS count_of
_orders,
               Extract (year FROM order_purchase_timestamp) AS year,
               Format_date("%B", order_purchase_timestamp)  AS month_na
me
       FROM   Ecommerce.orders
       WHERE  order_status NOT IN ( 'canceled', 'unavailable' )
       GROUP  BY year,
                 month_name)x
ORDER  BY count_of_orders DESC;
```

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS

Row	month_name	year	count_of_orders	percentage_of_orders
1	November	2017	7423	7.56
2	January	2018	7187	7.32
3	March	2018	7168	7.3
4	April	2018	6919	7.05
5	May	2018	6833	6.96
6	February	2018	6625	6.75
7	August	2018	6421	6.54
8	July	2018	6233	6.35
9	June	2018	6145	6.26
10	December	2017	5620	5.72

Reference:

https://www.independent.co.uk/news/world/black-friday-2017-brazil-shoppers-discount-sales-brazil-south-africa-a8073651.html

From the above result, we can illustrate below:

1) In 2017, November saw the highest sales, and in 2018, January, March, and April saw the highest sales.
2) From November 24 to November 28, 2017, sales in Brazil surged dramatically as a result of Black Friday and Thanksgiving.
3) Sales grew in January as a result of post-holiday sales, Christmas, and New Year's Eve.
4) Sales fall in February because consumers rein in their spending because they used to spend more from November to January.
5) Sales in March rise as a result of two significant occasions, International Women's Day and Consumers Day from 6.75% in February to 7.3% in March, sales grow.

BLACK FRIDAY WEEK SALES:

```
SELECT Ordered_day,
       week,
       month_name,
       year,
       count_of_orders,
       Round(( count_of_orders * 100 ) / ( Sum(count_of_orders)
                                             OVER () ), 2) AS
       percentage_of_orders,
       Delivered_in_day
FROM   (SELECT Count(order_id)                                  AS count_o
f_orders,
               Extract (year FROM order_purchase_timestamp) AS year,
               Format_date("%B", order_purchase_timestamp)  AS month_n
ame,
               Extract (day FROM order_purchase_timestamp)  AS Ordered
_day,
               Extract (week FROM order_purchase_timestamp) AS week,
               Round(Avg(Timestamp_diff(order_delivered_customer_date,
                         order_purchase_timestamp
                         , day)
                    ), 2)                                       AS Deliver
ed_in_day,
        FROM   Ecommerce.orders
        WHERE  order_status NOT IN ( 'canceled', 'unavailable' )
               AND Format_date("%B", order_purchase_timestamp) = 'Nove
mber'
        GROUP  BY year,
                  month_name,
                  week,
                  Ordered_day)x
ORDER  BY count_of_orders DESC;
```

Row	Ordered_day	week	month_name	year	count_of_orders	percentage_of_orders	Delivered_in_day
1	24	47	November	2017	1166	15.71	16.72
2	25	47	November	2017	499	6.72	18.96
3	27	48	November	2017	400	5.39	15.97
4	26	48	November	2017	387	5.21	16.08
5	28	48	November	2017	376	5.07	15.06

```sql
SELECT week,
       month_name,
       year,
       count_of_orders,
       Round(( count_of_orders * 100 ) / ( Sum(count_of_orders)
                                             OVER () ), 2) AS
       percentage_of_orders
FROM   (SELECT Count(order_id)                                  AS count_o
f_orders,
               Extract (year FROM order_purchase_timestamp) AS year,
               Format_date("%B", order_purchase_timestamp)  AS month_n
ame,
               Extract (week FROM order_purchase_timestamp) AS week
        FROM   Ecommerce.orders
        WHERE  order_status NOT IN ( 'canceled', 'unavailable' )
               AND Format_date("%B", order_purchase_timestamp) = 'Nove
mber'
        GROUP  BY year,
                  month_name,
                  week)x
ORDER  BY count_of_orders DESC;
```

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS

Row	week	month_name	year	count_of_orders	percentage_of_orders
1	47	November	2017	2744	36.97

Approximately 37% of sales in 2017 were generated during the Black Friday week in November.

Top products sold in Black Friday week in 2017:

```sql
SELECT product_category,
       Count(o.order_id) AS Ordercount
FROM   `Ecommerce.order_items` oi
       JOIN `Ecommerce.products` p
         ON oi.product_id = p.product_id
       JOIN `Ecommerce.orders` o
         ON oi.order_id = o.order_id
WHERE  order_status = 'delivered'
       AND Extract (month FROM order_purchase_timestamp) = 11
       AND Extract (year FROM order_purchase_timestamp) = 2017
GROUP  BY product_category
ORDER  BY ordercount DESC;
```

Row	product_category	Ordercount
1	bed table bath	961
2	Furniture Decoration	767
3	sport leisure	597
4	HEALTH BEAUTY	573
5	Garden tools	547
6	computer accessories	512
7	toys	485
8	Watches present	463
9	housewares	412
10	telephony	371

We may see from the above result that a range of goods were sold on which household goods made up the majority.

SALES ACCORDING TO MONTHS

```
SELECT month_name,
       count_of_orders,
       Round(( count_of_orders * 100 ) / ( Sum(count_of_orders)
                                             OVER () ), 2) AS
       percentage_of_orders
FROM  (SELECT Count(order_id)                      AS count_of_orders,
              Format_date("%B", order_purchase_timestamp) AS month_name
       FROM   Ecommerce.orders
       WHERE  order_status NOT IN ( 'canceled', 'unavailable' )
       GROUP  BY month_name)x
ORDER  BY count_of_orders DESC;
```

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS

Row	month_name	count_of_or...	percentage_...
1	August	10693	10.89
2	May	10473	10.66
3	July	10179	10.36
4	March	9785	9.96
5	June	9350	9.52
6	April	9296	9.47
7	February	8343	8.5
8	January	7974	8.12
9	November	7423	7.56
10	December	5621	5.72
11	October	4840	4.93

According to the results above, the month of August saw the most sales, followed by May and July.

Father's Day, which is celebrated in August in Brazil and is the most significant holiday after Black Friday and Christmas, and e-commerce provide significant discounts during the sale.

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

SALES ACCORDING TO TIME

```
SELECT Time,
       count_of_orders,
       Round(( count_of_orders * 100 ) / ( Sum(count_of_orders)
                                             OVER () ), 2) AS
       percentage_of_orders
FROM   (SELECT Count(order_id) AS count_of_orders,
               CASE
                 WHEN hour >= 4
                      AND hour < 7 THEN 'Dawn'
                 WHEN hour >= 7
                      AND hour < 12 THEN 'Morning'
                 WHEN hour >= 12
                      AND hour < 17 THEN 'Afternoon'
                 ELSE 'Night'
               END              AS Time
        FROM   (SELECT order_id,
```

```
                    Extract (hour FROM order_purchase_timestamp) AS hour
                FROM    Ecommerce.orders)x
        GROUP  BY Time) x
ORDER  BY count_of_orders DESC;
```

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS

Row	Time	count_of_orders	percentage_of_orders
1	Night	44596	44.85
2	Afternoon	32211	32.39
3	Morning	21738	21.86
4	Dawn	896	0.9

Consumers prefer to buy more products in night followed by afternoon.

3. **Evolution of E-commerce orders in the Brazil region:**
    1. **Get month on month orders by region, states**

Reference For States and Region: https://brazil-help.com/brazilian_states.htm

SALES BY STATES

```
SELECT CASE
          WHEN customer_state = 'SE' THEN 'Sergipe'
          WHEN customer_state = 'AL' THEN 'Alagoas'
          WHEN customer_state = 'PI' THEN 'Piauí'
          WHEN customer_state = 'AP' THEN 'Amapá'
          WHEN customer_state = 'AM' THEN 'Amazonas'
          WHEN customer_state = 'RR' THEN 'Roraima'
          WHEN customer_state = 'AC' THEN 'Acre'
          WHEN customer_state = 'RO' THEN 'Rondônia'
          WHEN customer_state = 'TO' THEN 'Tocantins'
          WHEN customer_state = 'BA' THEN 'Bahia'
          WHEN customer_state = 'CE' THEN 'Ceará'
          WHEN customer_state = 'DF' THEN 'Distrito Federal'
          WHEN customer_state = 'ES' THEN 'Espírito Santo'
          WHEN customer_state = 'GO' THEN 'Goiás'
          WHEN customer_state = 'MA' THEN 'Maranhão'
          WHEN customer_state = 'MG' THEN 'Minas Gerais'
          WHEN customer_state = 'MS' THEN 'MatoGrosso do Sul'
          WHEN customer_state = 'MT' THEN 'MatoGrosso'
          WHEN customer_state = 'PA' THEN 'Pará'
          WHEN customer_state = 'PB' THEN 'Paraíba'
          WHEN customer_state = 'PE' THEN 'Pernambuco'
```

```
  WHEN customer_state = 'PR' THEN 'Paraná'
  WHEN customer_state = 'RJ' THEN 'Rio de Janeiro'
  WHEN customer_state = 'RN' THEN 'Rio Grande do Norte'
  WHEN customer_state = 'RS' THEN 'Rio Grande do Sul'
  WHEN customer_state = 'SC' THEN 'Santa Catarina'
  WHEN customer_state = 'SP' THEN 'São Paulo'
end      AS customer_state,
Sum(CASE
      WHEN month = 1 THEN 1
      ELSE 0
    end) AS Jan,
Sum(CASE
      WHEN month = 2 THEN 1
      ELSE 0
    end) AS Feb,
Sum(CASE
      WHEN month = 3 THEN 1
      ELSE 0
    end) AS March,
Sum(CASE
      WHEN month = 4 THEN 1
      ELSE 0
    end) AS April,
Sum(CASE
      WHEN month = 5 THEN 1
      ELSE 0
    end) AS May,
Sum(CASE
      WHEN month = 6 THEN 1
      ELSE 0
    end) AS June,
Sum(CASE
      WHEN month = 7 THEN 1
      ELSE 0
    end) AS July,
Sum(CASE
      WHEN month = 8 THEN 1
      ELSE 0
    end) AS Aug,
Sum(CASE
      WHEN month = 9 THEN 1
      ELSE 0
    end) AS Sept,
Sum(CASE
      WHEN month = 10 THEN 1
      ELSE 0
    end) AS Oct,
Sum(CASE
      WHEN month = 11 THEN 1
      ELSE 0
    end) AS Nov,
Sum(CASE
```

```
            WHEN month = 12 THEN 1
            ELSE 0
        end) AS DEC
FROM    (SELECT c.customer_state,
                Extract(month FROM order_purchase_timestamp) AS month
         FROM   `Ecommerce.orders` o
                JOIN `Ecommerce.customers` c
                  ON o.customer_id = c.customer_id)x
GROUP  BY customer_state
ORDER  BY jan DESC
```

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS

Row	customer_state	Jan	Feb	March	April	May	June	July	Aug	Sept	Oct	Nov	DEC
1	São Paulo	3351	3357	4047	3967	4632	4104	4381	4982	1648	1908	3012	2357
2	Rio de Janeiro	990	1176	1302	1172	1321	1128	1288	1307	612	725	1048	783
3	Minas Gerais	971	1063	1237	1061	1190	1080	1111	1177	511	600	943	691
4	Paraná	443	460	504	500	524	478	523	556	183	225	378	271
5	Rio Grande do Sul	427	473	569	488	559	526	565	599	279	276	422	283
6	Santa Catarina	345	316	362	351	379	321	356	365	157	189	303	193
7	Bahia	264	273	340	318	368	307	405	323	170	170	250	192
8	Goiás	164	176	199	177	226	184	192	213	88	117	157	127
9	Espírito Santo	159	186	182	188	228	204	206	200	93	104	170	113
10	Distrito Federal	151	196	207	183	208	220	243	232	97	104	168	131

Good number of sales is from state São Paulo, Rio de Janeiro and Minas which are the most popular states in Brazil in terms of GDP, industrial and population perspective.

SALES BY REGION

Reference For Sates and Region: https://brazil-help.com/brazilian_states.htm

```
SELECT Region,
       Sum(CASE
             WHEN month = 1 THEN 1
             ELSE 0
           end) AS Jan,
       Sum(CASE
             WHEN month = 2 THEN 1
             ELSE 0
           end) AS Feb,
       Sum(CASE
             WHEN month = 3 THEN 1
             ELSE 0
           end) AS March,
       Sum(CASE
             WHEN month = 4 THEN 1
```

```
            ELSE 0
          end) AS April,
        Sum(CASE
            WHEN month = 5 THEN 1
            ELSE 0
          end) AS May,
        Sum(CASE
            WHEN month = 6 THEN 1
            ELSE 0
          end) AS June,
        Sum(CASE
            WHEN month = 7 THEN 1
            ELSE 0
          end) AS July,
        Sum(CASE
            WHEN month = 8 THEN 1
            ELSE 0
          end) AS Aug,
        Sum(CASE
            WHEN month = 9 THEN 1
            ELSE 0
          end) AS Sept,
        Sum(CASE
            WHEN month = 10 THEN 1
            ELSE 0
          end) AS Oct,
        Sum(CASE
            WHEN month = 11 THEN 1
            ELSE 0
          end) AS Nov,
        Sum(CASE
            WHEN month = 12 THEN 1
            ELSE 0
          end) AS DEC
FROM   (SELECT CASE
               WHEN customer_state IN ( 'SE', 'RN', 'PI', 'PE', 'PB',
                                        'MA', 'CE', 'BA', 'AL' ) THEN
                 'Northeast'
               WHEN customer_state IN ( 'AC', 'AP', 'AM', 'PA',
                                        'RO', 'RR', 'TO' ) THEN 'North'
               WHEN customer_state IN ( 'DF', 'GO', 'MT', 'MS' ) THEN
                 'Central-West'
               WHEN customer_state IN ( 'ES', 'MG', 'RJ', 'SP' ) THEN
                 'Southeast'
               WHEN customer_state IN ( 'SC', 'RS', 'PR' ) THEN 'South
'
               end                                                AS Region,
               Extract(month FROM order_purchase_timestamp) AS month
        FROM   `Ecommerce.orders` o
               JOIN `Ecommerce.customers` c
                 ON o.customer_id = c.customer_id)x
```

```
GROUP  BY Region
ORDER  BY jan DESC
```

Row	Region	Jan	Feb	March	April	May	June	July	Aug	Sept	Oct	Nov	DEC
1	Southeast	5471	5782	6768	6388	7371	6516	6986	7666	2864	3337	5173	3944
2	South	1215	1249	1435	1339	1462	1325	1444	1520	619	690	1103	747
3	Northeast	744	777	934	909	950	841	1103	899	477	521	698	541
4	Central-West	482	531	556	510	612	563	594	582	253	310	445	344
5	North	157	169	200	197	178	167	191	176	92	101	125	98

The Southeast Region, which comprises states like Espirito Santo, Minas Gerais, Rio de Janeiro, and So Paulo, accounts for the majority of sales.

2. How are customers distributed in Brazil

```
SELECT Count(customer_id) AS Total_Customers,
       CASE
         WHEN customer_state = 'SE' THEN 'Sergipe'
         WHEN customer_state = 'AL' THEN 'Alagoas'
         WHEN customer_state = 'PI' THEN 'Piauí'
         WHEN customer_state = 'AP' THEN 'Amapá'
         WHEN customer_state = 'AM' THEN 'Amazonas'
         WHEN customer_state = 'RR' THEN 'Roraima'
         WHEN customer_state = 'AC' THEN 'Acre'
         WHEN customer_state = 'RO' THEN 'Rondônia'
         WHEN customer_state = 'TO' THEN 'Tocantins'
         WHEN customer_state = 'BA' THEN 'Bahia'
         WHEN customer_state = 'CE' THEN 'Ceará'
         WHEN customer_state = 'DF' THEN 'Distrito Federal'
         WHEN customer_state = 'ES' THEN 'Espírito Santo'
         WHEN customer_state = 'GO' THEN 'Goiás'
         WHEN customer_state = 'MA' THEN 'Maranhão'
         WHEN customer_state = 'MG' THEN 'Minas Gerais'
         WHEN customer_state = 'MS' THEN 'MatoGrosso do Sul'
         WHEN customer_state = 'MT' THEN 'MatoGrosso'
         WHEN customer_state = 'PA' THEN 'Pará'
         WHEN customer_state = 'PB' THEN 'Paraíba'
         WHEN customer_state = 'PE' THEN 'Pernambuco'
         WHEN customer_state = 'PR' THEN 'Paraná'
         WHEN customer_state = 'RJ' THEN 'Rio de Janeiro'
         WHEN customer_state = 'RN' THEN 'Rio Grande do Norte'
         WHEN customer_state = 'RS' THEN 'Rio Grande do Sul'
         WHEN customer_state = 'SC' THEN 'Santa Catarina'
         WHEN customer_state = 'SP' THEN 'São Paulo'
       end                  AS customer_state,
       customer_city
```

```
FROM    `Ecommerce.customers`
GROUP   BY customer_state,
           customer_city
ORDER   BY Total_Customers DESC
```

Row	Total_Customers	customer_state	customer_city
1	15540	São Paulo	sao paulo
2	6882	Rio de Janeiro	rio de janeiro
3	2773	Minas Gerais	belo horizonte
4	2131	Distrito Federal	brasilia
5	1521	Paraná	curitiba
6	1444	São Paulo	campinas
7	1379	Rio Grande do Sul	porto alegre
8	1245	Bahia	salvador
9	1189	São Paulo	guarulhos
10	938	São Paulo	sao bernardo do campo

Maximum customers are from state São Paulo and sao paulo city and hence has a greater number of orders among other states and country which is depicted below.

TOTAL SALES ACCORDING TO STATE:

```
SELECT CASE
          WHEN customer_state = 'SE' THEN 'Sergipe'
          WHEN customer_state = 'AL' THEN 'Alagoas'
          WHEN customer_state = 'PI' THEN 'Piauí'
          WHEN customer_state = 'AP' THEN 'Amapá'
          WHEN customer_state = 'AM' THEN 'Amazonas'
          WHEN customer_state = 'RR' THEN 'Roraima'
          WHEN customer_state = 'AC' THEN 'Acre'
          WHEN customer_state = 'RO' THEN 'Rondônia'
          WHEN customer_state = 'TO' THEN 'Tocantins'
          WHEN customer_state = 'BA' THEN 'Bahia'
          WHEN customer_state = 'CE' THEN 'Ceará'
          WHEN customer_state = 'DF' THEN 'Distrito Federal'
          WHEN customer_state = 'ES' THEN 'Espírito Santo'
          WHEN customer_state = 'GO' THEN 'Goiás'
          WHEN customer_state = 'MA' THEN 'Maranhão'
          WHEN customer_state = 'MG' THEN 'Minas Gerais'
          WHEN customer_state = 'MS' THEN 'MatoGrosso do Sul'
          WHEN customer_state = 'MT' THEN 'MatoGrosso'
          WHEN customer_state = 'PA' THEN 'Pará'
          WHEN customer_state = 'PB' THEN 'Paraíba'
```

```
          WHEN customer_state = 'PE' THEN 'Pernambuco'
          WHEN customer_state = 'PR' THEN 'Paraná'
          WHEN customer_state = 'RJ' THEN 'Rio de Janeiro'
          WHEN customer_state = 'RN' THEN 'Rio Grande do Norte'
          WHEN customer_state = 'RS' THEN 'Rio Grande do Sul'
          WHEN customer_state = 'SC' THEN 'Santa Catarina'
          WHEN customer_state = 'SP' THEN 'São Paulo'
        end AS customer_state,
        Total_orders
FROM   (SELECT customer_state,
               Count(o.order_id) AS Total_orders
        FROM   `Ecommerce.orders` o
               JOIN `Ecommerce.customers` c
                 ON o.customer_id = c.customer_id
                    AND order_status = 'delivered'
        GROUP  BY customer_state)x
ORDER  BY Total_orders DESC
```

JOB INFORMATION | RESULTS

Row	customer_state	Total_orders
1	São Paulo	40501
2	Rio de Janeiro	12350
3	Minas Gerais	11354
4	Rio Grande do Sul	5345
5	Paraná	4923
6	Santa Catarina	3546
7	Bahia	3256
8	Distrito Federal	2080
9	Espírito Santo	1995
10	Goiás	1957

TOTAL SALES ACCORDING TO CITY:

```
SELECT customer_city,
       Count(o.order_id) AS Total_orders
FROM   `Ecommerce.orders` o
```

```
        JOIN `Ecommerce.customers` c
          ON o.customer_id = c.customer_id
             AND order_status = 'delivered'
GROUP  BY customer_city
ORDER  BY Total_orders DESC
```

JOB INFORMATION | RESULTS | JSON

Row	customer_city	Total_orders
1	sao paulo	15045
2	rio de janeiro	6601
3	belo horizonte	2697
4	brasilia	2071
5	curitiba	1489
6	campinas	1406
7	porto alegre	1342
8	salvador	1188
9	guarulhos	1144
10	sao bernardo do campo	911

4. **Impact on Economy: Analyze the money movemented by e-commerce by looking at order prices, freight and others.**
    1. **Get % increase in cost of orders from 2017 to 2018 (include months between Jan to Aug only)**

```
WITH ctecost AS
(
       SELECT o.order_id,
              (price+freight_value)                          AS price,
              Extract(MONTH FROM order_purchase_timestamp)  AS MONTH,
              Extract(YEAR FROM order_purchase_timestamp)   AS YEAR
       FROM    `Ecommerce.orders` o
       JOIN    `Ecommerce.order_items` oi
       ON      o.order_id=oi.order_id
       AND     order_status='delivered')
SELECT
       CASE
              WHEN MONTH=1 THEN 'Jan'
```

```
                WHEN MONTH=2 THEN 'Feb'
                WHEN MONTH=3 THEN 'Mar'
                WHEN MONTH=4 THEN 'April'
                WHEN MONTH=5 THEN 'May'
                WHEN MONTH=6 THEN 'June'
                WHEN MONTH=7 THEN 'July'
                WHEN MONTH=8 THEN 'Aug'
        END AS MONTH,
        Order_Count_in_2017,
        Order_Count_in_2018,
        Total_Cost_in_2017,
        Total_Cost_in_2018,
        ROUND((Total_Cost_in_2018-
Total_Cost_in_2017)*100/Total_Cost_in_2017,2) AS percent_increase
FROM  (
                SELECT   MONTH,
                         COUNT(
                         CASE
                                  WHEN YEAR=2017 THEN order_id
                         END) AS Order_Count_in_2017,
                         COUNT(
                         CASE
                                  WHEN YEAR=2018 THEN order_id
                         END) AS Order_Count_in_2018,
                         SUM(
                         CASE
                                  WHEN YEAR=2017 THEN ROUND(price,0)
                                  ELSE 0
                         END) AS Total_Cost_in_2017,
                         SUM(
                         CASE
                                  WHEN YEAR=2018 THEN ROUND(price,0)
                                  ELSE 0
                         END) AS Total_Cost_in_2018
                FROM     ctecost
                WHERE    MONTH NOT IN (9,10,11,12)
                GROUP BY MONTH
                ORDER BY MONTH)x
```

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS

Row	Month	Order_Count_in_2017	Order_Count_in_2018	Total_Cost_in_2017	Total_Cost_in_2018	percent_increase
1	Jan	913	8037	127524.0	1078173.0	745.47
2	Feb	1858	7518	271295.0	966394.0	256.22
3	Mar	2897	8017	414440.0	1120499.0	170.36
4	April	2569	7827	390872.0	1132732.0	189.8
5	May	4004	7810	566939.0	1128669.0	99.08
6	June	3489	7010	490168.0	1011824.0	106.42
7	July	4416	6963	566426.0	1027696.0	81.44
8	Aug	4797	7142	645993.0	985312.0	52.53

We can see from the following data that monthly sales grew from 2017 to 2018. For instance, sales increase by 8 times in January, 2.5 times in February, and so on.

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

```
SELECT CASE
         WHEN c.customer_state = 'SE' THEN 'Sergipe'
         WHEN c.customer_state = 'AL' THEN 'Alagoas'
         WHEN c.customer_state = 'PI' THEN 'Piauí'
         WHEN c.customer_state = 'AP' THEN 'Amapá'
         WHEN c.customer_state = 'AM' THEN 'Amazonas'
         WHEN c.customer_state = 'RR' THEN 'Roraima'
         WHEN c.customer_state = 'AC' THEN 'Acre'
         WHEN c.customer_state = 'RO' THEN 'Rondônia'
         WHEN c.customer_state = 'TO' THEN 'Tocantins'
         WHEN c.customer_state = 'BA' THEN 'Bahia'
         WHEN c.customer_state = 'CE' THEN 'Ceará'
         WHEN c.customer_state = 'DF' THEN 'Distrito Federal'
         WHEN c.customer_state = 'ES' THEN 'Espírito Santo'
         WHEN c.customer_state = 'GO' THEN 'Goiás'
         WHEN c.customer_state = 'MA' THEN 'Maranhão'
         WHEN c.customer_state = 'MG' THEN 'Minas Gerais'
         WHEN c.customer_state = 'MS' THEN 'MatoGrosso do Sul'
         WHEN c.customer_state = 'MT' THEN 'MatoGrosso'
         WHEN c.customer_state = 'PA' THEN 'Pará'
         WHEN c.customer_state = 'PB' THEN 'Paraíba'
         WHEN c.customer_state = 'PE' THEN 'Pernambuco'
         WHEN c.customer_state = 'PR' THEN 'Paraná'
         WHEN c.customer_state = 'RJ' THEN 'Rio de Janeiro'
         WHEN c.customer_state = 'RN' THEN 'Rio Grande do Norte'
         WHEN c.customer_state = 'RS' THEN 'Rio Grande do Sul'
         WHEN c.customer_state = 'SC' THEN 'Santa Catarina'
         WHEN c.customer_state = 'SP' THEN 'São Paulo'
       end                                AS customer_state,
       Round(Avg(oi.price), 2)            AS Mean_of_Price,
       Round(Avg(oi.freight_value), 2)    AS Mean_of_Freight_Value,
       Round(Sum(oi.price), 2)            AS Sum_of_Price,
```

```
        Round(Sum(oi.freight_value), 2) AS Sum_of_Freight_Value
FROM    `Ecommerce.orders` o
        JOIN `Ecommerce.order_items` oi
          ON o.order_id = oi.order_id
        JOIN `Ecommerce.customers` c
          ON o.customer_id = c.customer_id
             AND order_status = 'delivered'
GROUP   BY customer_state
ORDER   BY Mean_of_Price DESC
```

Row	customer_state	Mean_of_Price	Mean_of_Freight_Value	Sum_of_Price	Sum_of_Freight_Value
1	Paraíba	192.13	43.09	112586.82	25251.73
2	Alagoas	184.67	35.87	78855.72	15316.77
3	Acre	175.07	40.05	15930.97	3644.36
4	Rondônia	167.34	41.33	45682.76	11283.24
5	Pará	165.53	35.63	174470.59	37552.98
6	Amapá	165.12	34.16	13374.81	2767.0
7	Piauí	161.99	39.12	84721.0	20457.19
8	Rio Grande do Norte	157.59	35.72	82105.66	18609.12
9	Tocantins	156.14	37.44	48402.51	11604.86
10	Ceará	154.11	32.73	219757.38	46679.39

Paraíba and Alagoas states has highest avg price and freight value compared to other states.

5. **Analysis on sales, freight and delivery time**
    1. **Calculate days between purchasing, delivering and estimated delivery**

Time to Deliver in Ascending Order:

```
SELECT c.customer_id,
       CASE
         WHEN customer_state = 'SE' THEN 'Sergipe'
         WHEN customer_state = 'AL' THEN 'Alagoas'
         WHEN customer_state = 'PI' THEN 'Piauí'
         WHEN customer_state = 'AP' THEN 'Amapá'
         WHEN customer_state = 'AM' THEN 'Amazonas'
         WHEN customer_state = 'RR' THEN 'Roraima'
         WHEN customer_state = 'AC' THEN 'Acre'
         WHEN customer_state = 'RO' THEN 'Rondônia'
         WHEN customer_state = 'TO' THEN 'Tocantins'
         WHEN customer_state = 'BA' THEN 'Bahia'
         WHEN customer_state = 'CE' THEN 'Ceará'
         WHEN customer_state = 'DF' THEN 'Distrito Federal'
         WHEN customer_state = 'ES' THEN 'Espírito Santo'
         WHEN customer_state = 'GO' THEN 'Goiás'
         WHEN customer_state = 'MA' THEN 'Maranhão'
```

```
          WHEN customer_state = 'MG' THEN 'Minas Gerais'
          WHEN customer_state = 'MS' THEN 'MatoGrosso do Sul'
          WHEN customer_state = 'MT' THEN 'MatoGrosso'
          WHEN customer_state = 'PA' THEN 'Pará'
          WHEN customer_state = 'PB' THEN 'Paraíba'
          WHEN customer_state = 'PE' THEN 'Pernambuco'
          WHEN customer_state = 'PR' THEN 'Paraná'
          WHEN customer_state = 'RJ' THEN 'Rio de Janeiro'
          WHEN customer_state = 'RN' THEN 'Rio Grande do Norte'
          WHEN customer_state = 'RS' THEN 'Rio Grande do Sul'
          WHEN customer_state = 'SC' THEN 'Santa Catarina'
          WHEN customer_state = 'SP' THEN 'São Paulo'
        end
        AS customer_state,
        order_estimated_delivery_date
        AS EstimateDeliveryDay,
        order_delivered_customer_date
        AS DeliveryDay,
        order_purchase_timestamp
        AS OrderedDay,
        Timestamp_diff(order_delivered_customer_date, order_purchase_ti
mestamp,
        day) AS
        Delivered_in_Day,
        Timestamp_diff(order_estimated_delivery_date, order_purchase_ti
mestamp,
        day) AS
        Diff_ExpectedDelivery_and_Purchase,
        Timestamp_diff(order_estimated_delivery_date,
        order_delivered_customer_date, day
        )
        AS Diff_ExpectedDelivery_and_Delivery
FROM    `Ecommerce.orders` o
        JOIN `Ecommerce.customers` c
          ON o.customer_id = c.customer_id
WHERE   order_status = 'delivered'
        AND order_delivered_customer_date IS NOT NULL
ORDER   BY Delivered_in_Day
```

R...	customer_id	customer_state	EstimateDeliveryDay	DeliveryDay	OrderedDay	Delivered_in_Day	Diff_ExpectedDelivery_and_Purchase	Diff_ExpectedDelivery_and...
1	198f511b5a75b...	São Paulo	2018-05-29 00:00:0...	2018-05-19 12:28:...	2018-05-18 15:03:...	0	10	9
2	118295a853acb...	São Paulo	2017-06-27 00:00:0...	2017-06-01 08:34:...	2017-05-31 11:11:...	0	26	25
3	922a46283625...	São Paulo	2017-06-19 00:00:0...	2017-05-30 08:06:...	2017-05-29 13:21:...	0	20	19
4	d23df2c6c3e51...	São Paulo	2017-11-29 00:00:0...	2017-11-17 13:49:...	2017-11-16 13:54:...	0	12	11
5	b19da0df0271e...	Rio de Janeiro	2017-06-30 00:00:0...	2017-06-19 21:07:...	2017-06-19 08:19:...	0	10	10
6	344423c2e26d4...	São Paulo	2017-05-24 00:00:0...	2017-05-16 10:21:...	2017-05-15 11:50:...	0	8	7
7	331d79b67223...	São Paulo	2018-06-28 00:00:0...	2018-06-19 12:43:...	2018-06-18 12:59:...	0	9	8
8	18c934f4cdc99...	São Paulo	2018-02-20 00:00:0...	2018-02-03 15:05:...	2018-02-02 15:26:...	0	17	16
9	ff58662c328f81...	Bahia	2018-07-25 00:00:0...	2018-06-27 17:31:...	2018-06-26 20:48:...	0	28	27
10	c5e200d485ae3...	São Paulo	2018-07-12 00:00:0...	2018-06-29 14:12:...	2018-06-28 14:34:...	0	13	12

There are a few orders that were delivered faster than expected—often to clients from São Paulo and Rio de Janeiro state—when the expected delivery time was more than 10 days.

Time to Deliver in Descending Order:

```
SELECT c.customer_id,
       CASE
         WHEN customer_state = 'SE' THEN 'Sergipe'
         WHEN customer_state = 'AL' THEN 'Alagoas'
         WHEN customer_state = 'PI' THEN 'Piauí'
         WHEN customer_state = 'AP' THEN 'Amapá'
         WHEN customer_state = 'AM' THEN 'Amazonas'
         WHEN customer_state = 'RR' THEN 'Roraima'
         WHEN customer_state = 'AC' THEN 'Acre'
         WHEN customer_state = 'RO' THEN 'Rondônia'
         WHEN customer_state = 'TO' THEN 'Tocantins'
         WHEN customer_state = 'BA' THEN 'Bahia'
         WHEN customer_state = 'CE' THEN 'Ceará'
         WHEN customer_state = 'DF' THEN 'Distrito Federal'
         WHEN customer_state = 'ES' THEN 'Espírito Santo'
         WHEN customer_state = 'GO' THEN 'Goiás'
         WHEN customer_state = 'MA' THEN 'Maranhão'
         WHEN customer_state = 'MG' THEN 'Minas Gerais'
         WHEN customer_state = 'MS' THEN 'MatoGrosso do Sul'
         WHEN customer_state = 'MT' THEN 'MatoGrosso'
         WHEN customer_state = 'PA' THEN 'Pará'
         WHEN customer_state = 'PB' THEN 'Paraíba'
         WHEN customer_state = 'PE' THEN 'Pernambuco'
         WHEN customer_state = 'PR' THEN 'Paraná'
         WHEN customer_state = 'RJ' THEN 'Rio de Janeiro'
         WHEN customer_state = 'RN' THEN 'Rio Grande do Norte'
         WHEN customer_state = 'RS' THEN 'Rio Grande do Sul'
         WHEN customer_state = 'SC' THEN 'Santa Catarina'
         WHEN customer_state = 'SP' THEN 'São Paulo'
       end
       AS customer_state,
       order_estimated_delivery_date
       AS EstimateDeliveryDay,
       order_delivered_customer_date
       AS DeliveryDay,
       order_purchase_timestamp
       AS OrderedDay,
       Timestamp_diff(order_delivered_customer_date, order_purchase_ti
mestamp,
       day) AS
       Delivered_in_Day,
       Timestamp_diff(order_estimated_delivery_date, order_purchase_ti
mestamp,
```

```
        day) AS
        Diff_ExpectedDelivery_and_Purchase,
        Timestamp_diff(order_estimated_delivery_date,
        order_delivered_customer_date, day
        )
        AS Diff_ExpectedDelivery_and_Delivery
FROM    `Ecommerce.orders` o
        JOIN `Ecommerce.customers` c
          ON o.customer_id = c.customer_id
WHERE   order_status = 'delivered'
        AND order_delivered_customer_date IS NOT NULL
ORDER   BY Delivered_in_Day DESC
```

Row	customer_id	customer_state	EstimateDeliveryDay	DeliveryDay	OrderedDay	Delivered_in_Day	Diff_ExpectedDelivery_and_Purchase	Diff_ExpectedDelivery_and_Delivery
1	75683a923310...	Espírito Santo	2017-03-22 00:00:...	2017-09-19...	2017-02-21 ...	209	28	-181
2	d306426abe5fc...	Rio de Janeiro	2018-03-15 00:00:...	2018-09-19...	2018-02-23 ...	208	19	-188
3	7815125148cfa...	Pará	2017-04-07 00:00:...	2017-09-19...	2017-03-07 ...	195	30	-165
4	1a8a4a30dc29...	Piauí	2017-04-11 00:00:...	2017-09-19...	2017-03-09 ...	194	32	-161
5	9cf2c3fa2632c...	Sergipe	2017-04-06 00:00:...	2017-09-19...	2017-03-08 ...	194	28	-166
6	217906bc11a3...	Piauí	2017-04-17 00:00:...	2017-09-19...	2017-03-08 ...	194	39	-155
7	cb2caaaead40...	São Paulo	2018-01-19 00:00:...	2018-07-13...	2018-01-03 ...	191	15	-175
8	65b14237885b...	São Paulo	2017-04-05 00:00:...	2017-09-19...	2017-03-13 ...	189	22	-167
9	8199345f57c6d...	Sergipe	2017-04-13 00:00:...	2017-09-19...	2017-03-15 ...	188	28	-159
10	9b39de85d94d...	Amapá	2017-04-28 00:00:...	2017-09-19...	2017-03-16 ...	187	42	-144

Occasionally, orders that were supposed to arrive in 20 to 30 days did so in excess of 100 days.

Avg Time to Deliver group by customer state:

```
SELECT CASE
         WHEN customer_state = 'SE' THEN 'Sergipe'
         WHEN customer_state = 'AL' THEN 'Alagoas'
         WHEN customer_state = 'PI' THEN 'Piauí'
         WHEN customer_state = 'AP' THEN 'Amapá'
         WHEN customer_state = 'AM' THEN 'Amazonas'
         WHEN customer_state = 'RR' THEN 'Roraima'
         WHEN customer_state = 'AC' THEN 'Acre'
         WHEN customer_state = 'RO' THEN 'Rondônia'
         WHEN customer_state = 'TO' THEN 'Tocantins'
         WHEN customer_state = 'BA' THEN 'Bahia'
         WHEN customer_state = 'CE' THEN 'Ceará'
         WHEN customer_state = 'DF' THEN 'Distrito Federal'
         WHEN customer_state = 'ES' THEN 'Espírito Santo'
         WHEN customer_state = 'GO' THEN 'Goiás'
```

```
      WHEN customer_state = 'MA' THEN 'Maranhão'
      WHEN customer_state = 'MG' THEN 'Minas Gerais'
      WHEN customer_state = 'MS' THEN 'MatoGrosso do Sul'
      WHEN customer_state = 'MT' THEN 'MatoGrosso'
      WHEN customer_state = 'PA' THEN 'Pará'
      WHEN customer_state = 'PB' THEN 'Paraíba'
      WHEN customer_state = 'PE' THEN 'Pernambuco'
      WHEN customer_state = 'PR' THEN 'Paraná'
      WHEN customer_state = 'RJ' THEN 'Rio de Janeiro'
      WHEN customer_state = 'RN' THEN 'Rio Grande do Norte'
      WHEN customer_state = 'RS' THEN 'Rio Grande do Sul'
      WHEN customer_state = 'SC' THEN 'Santa Catarina'
      WHEN customer_state = 'SP' THEN 'São Paulo'
    end              AS customer_state,
    Round(Avg(Timestamp_diff(order_delivered_customer_date,
              order_purchase_timestamp
              , day)
          ), 2)      AS Avg_Time_To_Deliver,
    Round(Avg(Timestamp_diff(order_estimated_delivery_date,
              order_purchase_timestamp
              , day)
          ), 2)      AS Avg_Diff_ExpectedDelivery_and_Purchase,
    Round(Avg(Timestamp_diff(order_estimated_delivery_date,
              order_delivered_customer_date,
                    day
              )), 2) AS Avg_Diff_ExpectedDelivery_and_Delivery
FROM  `Ecommerce.orders` o
    JOIN `Ecommerce.customers` c
      ON o.customer_id = c.customer_id
WHERE order_status = 'delivered'
GROUP BY customer_state
ORDER BY Avg_Time_To_Deliver
```

Row	customer_state	Avg_Time_To_Deliver	Avg_Diff_ExpectedDelivery_and_Purchase	Avg_Diff_ExpectedDelivery_and_Delivery
1	São Paulo	8.3	18.78	10.13
2	Paraná	11.53	24.25	12.36
3	Minas Gerais	11.54	24.19	12.3
4	Distrito Federal	12.51	23.95	11.12
5	Santa Catarina	14.48	25.41	10.6
6	Rio Grande do Sul	14.82	28.16	12.98
7	Rio de Janeiro	14.85	26.0	10.9
8	Goiás	15.15	26.72	11.27
9	MatoGrosso do Sul	15.19	25.6	10.17
10	Espírito Santo	15.33	25.22	9.62

Customers from the Brazilian states of São Paulo, Paraná, Minas Gerais, and Rio de Janeiro see shorter average delivery times.

Avg Time to Deliver group by seller state:

```
SELECT CASE
         WHEN seller_state = 'SE' THEN 'Sergipe'
         WHEN seller_state = 'AL' THEN 'Alagoas'
         WHEN seller_state = 'PI' THEN 'Piauí'
         WHEN seller_state = 'AP' THEN 'Amapá'
         WHEN seller_state = 'AM' THEN 'Amazonas'
         WHEN seller_state = 'RR' THEN 'Roraima'
         WHEN seller_state = 'AC' THEN 'Acre'
         WHEN seller_state = 'RO' THEN 'Rondônia'
         WHEN seller_state = 'TO' THEN 'Tocantins'
         WHEN seller_state = 'BA' THEN 'Bahia'
         WHEN seller_state = 'CE' THEN 'Ceará'
         WHEN seller_state = 'DF' THEN 'Distrito Federal'
         WHEN seller_state = 'ES' THEN 'Espírito Santo'
         WHEN seller_state = 'GO' THEN 'Goiás'
         WHEN seller_state = 'MA' THEN 'Maranhão'
         WHEN seller_state = 'MG' THEN 'Minas Gerais'
         WHEN seller_state = 'MS' THEN 'MatoGrosso do Sul'
         WHEN seller_state = 'MT' THEN 'MatoGrosso'
         WHEN seller_state = 'PA' THEN 'Pará'
         WHEN seller_state = 'PB' THEN 'Paraíba'
         WHEN seller_state = 'PE' THEN 'Pernambuco'
         WHEN seller_state = 'PR' THEN 'Paraná'
         WHEN seller_state = 'RJ' THEN 'Rio de Janeiro'
         WHEN seller_state = 'RN' THEN 'Rio Grande do Norte'
         WHEN seller_state = 'RS' THEN 'Rio Grande do Sul'
         WHEN seller_state = 'SC' THEN 'Santa Catarina'
         WHEN seller_state = 'SP' THEN 'São Paulo'
       end              AS seller_state,
       Round(Avg(Timestamp_diff(order_delivered_customer_date,
                 order_purchase_timestamp
                 , day)
             ), 2)      AS Avg_Delivered_in_Day,
       Round(Avg(Timestamp_diff(order_estimated_delivery_date,
                 order_purchase_timestamp
                 , day)
             ), 2)      AS Avg_Diff_ExpectedDelivery_and_Purchase,
       Round(Avg(Timestamp_diff(order_estimated_delivery_date,
                 order_delivered_customer_date,
                       day
                 )), 2) AS Avg_Diff_ExpectedDelivery_and_Delivery
FROM   `Ecommerce.orders` o
       JOIN `Ecommerce.customers` c
         ON o.customer_id = c.customer_id
       JOIN `Ecommerce.order_items` oi
         ON oi.order_id = o.order_id
       JOIN `Ecommerce.sellers` s
         ON oi.seller_id = s.seller_id
WHERE  order_status = 'delivered'
```

```
        AND order_delivered_customer_date IS NOT NULL
GROUP   BY seller_state
ORDER   BY Avg_Delivered_in_Day
```

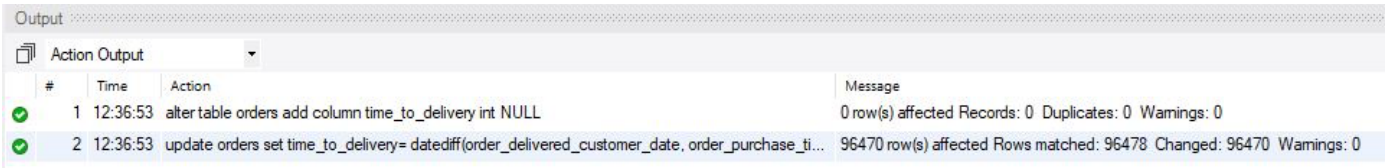
Row	seller_state	Avg_Delivered_in_Day	Avg_Diff_ExpectedDelivery_and_Purchase	Avg_Diff_ExpectedDelivery_and_Delivery
1	Rio Grande do Sul	11.09	26.82	15.37
2	Rio de Janeiro	11.56	23.42	11.56
3	São Paulo	11.81	22.51	10.38
4	MatoGrosso do Sul	11.9	28.68	16.46
5	Distrito Federal	12.09	24.62	12.25
6	Paraíba	12.16	31.16	18.84
7	Sergipe	12.2	28.9	16.3
8	Minas Gerais	12.33	25.22	12.53
9	Goiás	12.37	26.12	13.39
10	Espírito Santo	12.42	25.22	12.43

Customers whose sellers are from the states of Rio Grande do Sul, Rio Grande do Norte, São Paulo, and Mato Grosso do Sul see shorter delivery times.

2. **Create columns:**
   - **time_to_delivery = order_purchase_timestamp-order_delivered_customer_date**

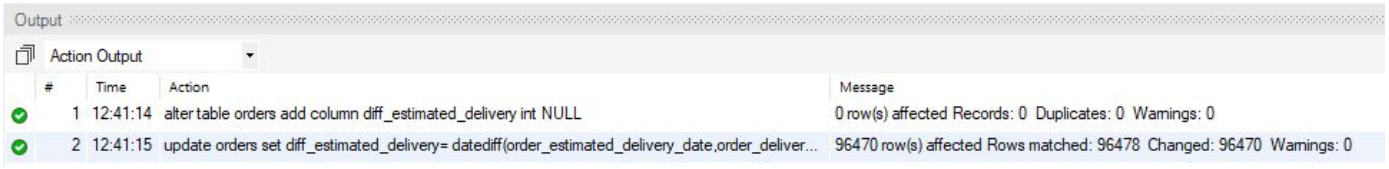
Using MySQL:

```
ALTER TABLE orders
  ADD COLUMN time_to_delivery INT NULL;

UPDATE orders
SET    time_to_delivery = Datediff(order_delivered_customer_date,
                                   order_purchase_timestamp)
WHERE  order_status = 'delivered';
```



- **diff_estimated_delivery = order_estimated_delivery_date-order_delivered_customer_date**

```
ALTER TABLE orders
  ADD COLUMN diff_estimated_delivery INT NULL;

UPDATE orders
SET    diff_estimated_delivery = Datediff(order_estimated_delivery_dat
e,
                                          order_delivered_customer_date)
WHERE  order_status = 'delivered';
```



Using CTE:

```
WITH cte_order_delivery AS
(
       SELECT customer_state,
              seller_state,
              Timestamp_diff(order_delivered_customer_date, order_purc
hase_timestamp , DAY)       AS time_to_delivery,
              Timestamp_diff(order_estimated_delivery_date, order_deli
vered_customer_date , DAY) AS diff_estimated_delivery,
              freight_value
       FROM   `Ecommerce.orders` o
       JOIN   `Ecommerce.order_items` oi
       ON     o.order_id = oi.order_id
       JOIN   `Ecommerce.customers` c
       ON     o.customer_id = c.customer_id
       JOIN   `Ecommerce.sellers` s
       ON     oi.seller_id=s.seller_id
       AND    order_status = 'delivered')


SELECT *
FROM   cte_order_delivery limit 2
```

Row	customer_state	seller_state	time_to_delivery	diff_estimated_delivery	freight_value
1	GO	SC	23	9	21.01
2	SP	SP	12	-5	9.3

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

```
WITH cte_order_delivery AS
(
       SELECT customer_state,
              seller_state,
              Timestamp_diff(order_delivered_customer_date, order_purc
hase_timestamp , DAY)      AS time_to_delivery,
              Timestamp_diff(order_estimated_delivery_date, order_deli
vered_customer_date , DAY) AS diff_estimated_delivery,
              freight_value
       FROM   `Ecommerce.orders` o
       JOIN   `Ecommerce.order_items` oi
       ON     o.order_id = oi.order_id
       JOIN   `Ecommerce.customers` c
       ON     o.customer_id = c.customer_id
       JOIN   `Ecommerce.sellers` s
       ON     oi.seller_id=s.seller_id
       AND    order_status = 'delivered')
SELECT
         CASE
                  WHEN customer_state = 'SE' THEN 'Sergipe'
                  WHEN customer_state = 'AL' THEN 'Alagoas'
                  WHEN customer_state = 'PI' THEN 'Piauí'
                  WHEN customer_state = 'AP' THEN 'Amapá'
                  WHEN customer_state = 'AM' THEN 'Amazonas'
                  WHEN customer_state = 'RR' THEN 'Roraima'
                  WHEN customer_state = 'AC' THEN 'Acre'
                  WHEN customer_state = 'RO' THEN 'Rondônia'
                  WHEN customer_state = 'TO' THEN 'Tocantins'
                  WHEN customer_state = 'BA' THEN 'Bahia'
                  WHEN customer_state = 'CE' THEN 'Ceará'
                  WHEN customer_state = 'DF' THEN 'Distrito Federal'
                  WHEN customer_state = 'ES' THEN 'Espírito Santo'
                  WHEN customer_state = 'GO' THEN 'Goiás'
                  WHEN customer_state = 'MA' THEN 'Maranhão'
                  WHEN customer_state = 'MG' THEN 'Minas Gerais'
                  WHEN customer_state = 'MS' THEN 'MatoGrosso do Sul'
                  WHEN customer_state = 'MT' THEN 'MatoGrosso'
                  WHEN customer_state = 'PA' THEN 'Pará'
                  WHEN customer_state = 'PB' THEN 'Paraíba'
                  WHEN customer_state = 'PE' THEN 'Pernambuco'
                  WHEN customer_state = 'PR' THEN 'Paraná'
                  WHEN customer_state = 'RJ' THEN 'Rio de Janeiro'
                  WHEN customer_state = 'RN' THEN 'Rio Grande do Norte
'
                  WHEN customer_state = 'RS' THEN 'Rio Grande do Sul'
                  WHEN customer_state = 'SC' THEN 'Santa Catarina'
```

```
                WHEN customer_state = 'SP' THEN 'São Paulo'
        END AS customer_state,
        CASE
                WHEN seller_state='SE' THEN 'Sergipe'
                WHEN seller_state='AL' THEN 'Alagoas'
                WHEN seller_state='PI' THEN 'Piauí'
                WHEN seller_state='AP' THEN 'Amapá'
                WHEN seller_state='AM' THEN 'Amazonas'
                WHEN seller_state='RR' THEN 'Roraima'
                WHEN seller_state='AC' THEN 'Acre'
                WHEN seller_state='RO' THEN 'Rondônia'
                WHEN seller_state='TO' THEN 'Tocantins'
                WHEN seller_state='BA' THEN 'Bahia'
                WHEN seller_state='CE' THEN 'Ceará'
                WHEN seller_state='DF' THEN 'Distrito Federal'
                WHEN seller_state='ES' THEN 'Espírito Santo'
                WHEN seller_state='GO' THEN 'Goiás'
                WHEN seller_state='MA' THEN 'Maranhão'
                WHEN seller_state='MG' THEN 'Minas Gerais'
                WHEN seller_state='MS' THEN 'MatoGrosso do Sul'
                WHEN seller_state='MT' THEN 'MatoGrosso'
                WHEN seller_state='PA' THEN 'Pará'
                WHEN seller_state='PB' THEN 'Paraíba'
                WHEN seller_state='PE' THEN 'Pernambuco'
                WHEN seller_state='PR' THEN 'Paraná'
                WHEN seller_state='RJ' THEN 'Rio de Janeiro'
                WHEN seller_state='RN' THEN 'Rio Grande do Norte'
                WHEN seller_state='RS' THEN 'Rio Grande do Sul'
                WHEN seller_state='SC' THEN 'Santa Catarina'
                WHEN seller_state='SP' THEN 'São Paulo'
        END                                     AS seller_state,
        ROUND(AVG(freight_value),2)             AS Mean_Freight_Value,
        ROUND(AVG(time_to_delivery),2)          AS Mean_Delivery_Time,
        ROUND(AVG(diff_estimated_delivery),2) AS Mean_Diff_EstimatedD
elivery_Time
FROM    cte_order_delivery
GROUP BY customer_state,
        seller_state
ORDER BY Mean_Freight_Value
```

Row	customer_state	seller_state	Mean_Freight_Value	Mean_Delivery_Time	Mean_Diff_EstimatedDelivery_Time
1	Rio Grande do Norte	Rio Grande do Norte	8.33	4.38	14.67
2	Distrito Federal	Distrito Federal	9.01	5.62	6.67
3	Paraíba	Paraíba	9.63	8.0	17.0
4	Maranhão	Maranhão	11.2	10.07	5.6
5	Goiás	Goiás	11.85	4.95	8.26
6	Rio de Janeiro	Rio de Janeiro	12.45	6.12	12.24
7	Pernambuco	Pernambuco	12.55	6.46	10.29
8	São Paulo	São Paulo	13.2	7.46	9.29
9	Alagoas	Pernambuco	14.56	16.44	8.33
10	Distrito Federal	Goiás	15.33	8.79	13.24

States like Rio Grande do Norte , Distrito Federal ,São Paulo, Paraná, Minas Gerais, Rio de Janeiro have less mean freight value and delivery time compared to other States.

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

Top 5 States with Highest Mean Freight Value in Descending Order:

```
SELECT CASE
          WHEN customer_state = 'SE' THEN 'Sergipe'
          WHEN customer_state = 'AL' THEN 'Alagoas'
          WHEN customer_state = 'PI' THEN 'Piauí'
          WHEN customer_state = 'AP' THEN 'Amapá'
          WHEN customer_state = 'AM' THEN 'Amazonas'
          WHEN customer_state = 'RR' THEN 'Roraima'
          WHEN customer_state = 'AC' THEN 'Acre'
          WHEN customer_state = 'RO' THEN 'Rondônia'
          WHEN customer_state = 'TO' THEN 'Tocantins'
          WHEN customer_state = 'BA' THEN 'Bahia'
          WHEN customer_state = 'CE' THEN 'Ceará'
          WHEN customer_state = 'DF' THEN 'Distrito Federal'
          WHEN customer_state = 'ES' THEN 'Espírito Santo'
          WHEN customer_state = 'GO' THEN 'Goiás'
          WHEN customer_state = 'MA' THEN 'Maranhão'
          WHEN customer_state = 'MG' THEN 'Minas Gerais'
          WHEN customer_state = 'MS' THEN 'MatoGrosso do Sul'
          WHEN customer_state = 'MT' THEN 'MatoGrosso'
          WHEN customer_state = 'PA' THEN 'Pará'
          WHEN customer_state = 'PB' THEN 'Paraíba'
          WHEN customer_state = 'PE' THEN 'Pernambuco'
          WHEN customer_state = 'PR' THEN 'Paraná'
          WHEN customer_state = 'RJ' THEN 'Rio de Janeiro'
          WHEN customer_state = 'RN' THEN 'Rio Grande do Norte'
          WHEN customer_state = 'RS' THEN 'Rio Grande do Sul'
          WHEN customer_state = 'SC' THEN 'Santa Catarina'
          WHEN customer_state = 'SP' THEN 'São Paulo'
       end                           AS customer_state,
       CASE
          WHEN seller_state = 'SE' THEN 'Sergipe'
          WHEN seller_state = 'AL' THEN 'Alagoas'
          WHEN seller_state = 'PI' THEN 'Piauí'
          WHEN seller_state = 'AP' THEN 'Amapá'
          WHEN seller_state = 'AM' THEN 'Amazonas'
          WHEN seller_state = 'RR' THEN 'Roraima'
```

```
          WHEN seller_state = 'AC' THEN 'Acre'
          WHEN seller_state = 'RO' THEN 'Rondônia'
          WHEN seller_state = 'TO' THEN 'Tocantins'
          WHEN seller_state = 'BA' THEN 'Bahia'
          WHEN seller_state = 'CE' THEN 'Ceará'
          WHEN seller_state = 'DF' THEN 'Distrito Federal'
          WHEN seller_state = 'ES' THEN 'Espírito Santo'
          WHEN seller_state = 'GO' THEN 'Goiás'
          WHEN seller_state = 'MA' THEN 'Maranhão'
          WHEN seller_state = 'MG' THEN 'Minas Gerais'
          WHEN seller_state = 'MS' THEN 'MatoGrosso do Sul'
          WHEN seller_state = 'MT' THEN 'MatoGrosso'
          WHEN seller_state = 'PA' THEN 'Pará'
          WHEN seller_state = 'PB' THEN 'Paraíba'
          WHEN seller_state = 'PE' THEN 'Pernambuco'
          WHEN seller_state = 'PR' THEN 'Paraná'
          WHEN seller_state = 'RJ' THEN 'Rio de Janeiro'
          WHEN seller_state = 'RN' THEN 'Rio Grande do Norte'
          WHEN seller_state = 'RS' THEN 'Rio Grande do Sul'
          WHEN seller_state = 'SC' THEN 'Santa Catarina'
          WHEN seller_state = 'SP' THEN 'São Paulo'
        end                              AS seller_state,
        Round(Avg(freight_value), 2) AS Mean_Freight_value
FROM    `Ecommerce.orders` o
        JOIN `Ecommerce.order_items` oi
          ON o.order_id = oi.order_id
        JOIN `Ecommerce.customers` c
          ON o.customer_id = c.customer_id
        JOIN `Ecommerce.sellers` s
          ON oi.seller_id = s.seller_id
             AND order_status = 'delivered'
GROUP   BY customer_state,
           seller_state
ORDER   BY Mean_Freight_value DESC
LIMIT   5;
```

Row	customer_state	seller_state	Mean_Freight_value
1	Goiás	Ceará	144.29
2	Pará	Ceará	117.08
3	Rio Grande do Norte	Ceará	97.44
4	Pará	Espírito Santo	91.75
5	Rondônia	Espírito Santo	89.82

The average freight value is high when the buyer and seller are from the different state as the cost of transport and unloading goods increases.

Top 5 States with Lowest Mean Freight Value in Ascending Order:

```
SELECT CASE
         WHEN customer_state = 'SE' THEN 'Sergipe'
         WHEN customer_state = 'AL' THEN 'Alagoas'
         WHEN customer_state = 'PI' THEN 'Piauí'
         WHEN customer_state = 'AP' THEN 'Amapá'
         WHEN customer_state = 'AM' THEN 'Amazonas'
         WHEN customer_state = 'RR' THEN 'Roraima'
         WHEN customer_state = 'AC' THEN 'Acre'
         WHEN customer_state = 'RO' THEN 'Rondônia'
         WHEN customer_state = 'TO' THEN 'Tocantins'
         WHEN customer_state = 'BA' THEN 'Bahia'
         WHEN customer_state = 'CE' THEN 'Ceará'
         WHEN customer_state = 'DF' THEN 'Distrito Federal'
         WHEN customer_state = 'ES' THEN 'Espírito Santo'
         WHEN customer_state = 'GO' THEN 'Goiás'
         WHEN customer_state = 'MA' THEN 'Maranhão'
         WHEN customer_state = 'MG' THEN 'Minas Gerais'
         WHEN customer_state = 'MS' THEN 'MatoGrosso do Sul'
         WHEN customer_state = 'MT' THEN 'MatoGrosso'
         WHEN customer_state = 'PA' THEN 'Pará'
         WHEN customer_state = 'PB' THEN 'Paraíba'
         WHEN customer_state = 'PE' THEN 'Pernambuco'
         WHEN customer_state = 'PR' THEN 'Paraná'
         WHEN customer_state = 'RJ' THEN 'Rio de Janeiro'
         WHEN customer_state = 'RN' THEN 'Rio Grande do Norte'
         WHEN customer_state = 'RS' THEN 'Rio Grande do Sul'
         WHEN customer_state = 'SC' THEN 'Santa Catarina'
         WHEN customer_state = 'SP' THEN 'São Paulo'
       end                          AS customer_state,
       CASE
         WHEN seller_state = 'SE' THEN 'Sergipe'
         WHEN seller_state = 'AL' THEN 'Alagoas'
         WHEN seller_state = 'PI' THEN 'Piauí'
         WHEN seller_state = 'AP' THEN 'Amapá'
         WHEN seller_state = 'AM' THEN 'Amazonas'
         WHEN seller_state = 'RR' THEN 'Roraima'
         WHEN seller_state = 'AC' THEN 'Acre'
         WHEN seller_state = 'RO' THEN 'Rondônia'
         WHEN seller_state = 'TO' THEN 'Tocantins'
         WHEN seller_state = 'BA' THEN 'Bahia'
         WHEN seller_state = 'CE' THEN 'Ceará'
         WHEN seller_state = 'DF' THEN 'Distrito Federal'
         WHEN seller_state = 'ES' THEN 'Espírito Santo'
         WHEN seller_state = 'GO' THEN 'Goiás'
         WHEN seller_state = 'MA' THEN 'Maranhão'
         WHEN seller_state = 'MG' THEN 'Minas Gerais'
         WHEN seller_state = 'MS' THEN 'MatoGrosso do Sul'
         WHEN seller_state = 'MT' THEN 'MatoGrosso'
         WHEN seller_state = 'PA' THEN 'Pará'
```

```
          WHEN seller_state = 'PB' THEN 'Paraíba'
          WHEN seller_state = 'PE' THEN 'Pernambuco'
          WHEN seller_state = 'PR' THEN 'Paraná'
          WHEN seller_state = 'RJ' THEN 'Rio de Janeiro'
          WHEN seller_state = 'RN' THEN 'Rio Grande do Norte'
          WHEN seller_state = 'RS' THEN 'Rio Grande do Sul'
          WHEN seller_state = 'SC' THEN 'Santa Catarina'
          WHEN seller_state = 'SP' THEN 'São Paulo'
        end                            AS seller_state,
        Round(Avg(freight_value), 2) AS Mean_Freight_value
FROM    `Ecommerce.orders` o
        JOIN `Ecommerce.order_items` oi
          ON o.order_id = oi.order_id
        JOIN `Ecommerce.customers` c
          ON o.customer_id = c.customer_id
        JOIN `Ecommerce.sellers` s
          ON oi.seller_id = s.seller_id
             AND order_status = 'delivered'
GROUP   BY customer_state,
           seller_state
ORDER   BY Mean_Freight_value
LIMIT   5;
```

Row	customer_state	seller_state	Mean_Freight_value
1	Rio Grande do Norte	Rio Grande do Norte	8.33
2	Distrito Federal	Distrito Federal	9.01
3	Paraíba	Paraíba	9.63
4	Maranhão	Maranhão	11.2
5	Goiás	Goiás	11.85

The average freight value is the lowest when the buyer and seller are from the same state since it costs less to transport and unload the goods.

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

Top 5 States with Highest Mean Delivery Time:

```
SELECT CASE
          WHEN customer_state = 'SE' THEN 'Sergipe'
          WHEN customer_state = 'AL' THEN 'Alagoas'
          WHEN customer_state = 'PI' THEN 'Piauí'
          WHEN customer_state = 'AP' THEN 'Amapá'
          WHEN customer_state = 'AM' THEN 'Amazonas'
          WHEN customer_state = 'RR' THEN 'Roraima'
```

```
  WHEN customer_state = 'AC' THEN 'Acre'
  WHEN customer_state = 'RO' THEN 'Rondônia'
  WHEN customer_state = 'TO' THEN 'Tocantins'
  WHEN customer_state = 'BA' THEN 'Bahia'
  WHEN customer_state = 'CE' THEN 'Ceará'
  WHEN customer_state = 'DF' THEN 'Distrito Federal'
  WHEN customer_state = 'ES' THEN 'Espírito Santo'
  WHEN customer_state = 'GO' THEN 'Goiás'
  WHEN customer_state = 'MA' THEN 'Maranhão'
  WHEN customer_state = 'MG' THEN 'Minas Gerais'
  WHEN customer_state = 'MS' THEN 'MatoGrosso do Sul'
  WHEN customer_state = 'MT' THEN 'MatoGrosso'
  WHEN customer_state = 'PA' THEN 'Pará'
  WHEN customer_state = 'PB' THEN 'Paraíba'
  WHEN customer_state = 'PE' THEN 'Pernambuco'
  WHEN customer_state = 'PR' THEN 'Paraná'
  WHEN customer_state = 'RJ' THEN 'Rio de Janeiro'
  WHEN customer_state = 'RN' THEN 'Rio Grande do Norte'
  WHEN customer_state = 'RS' THEN 'Rio Grande do Sul'
  WHEN customer_state = 'SC' THEN 'Santa Catarina'
  WHEN customer_state = 'SP' THEN 'São Paulo'
end         AS customer_state,
CASE
  WHEN seller_state = 'SE' THEN 'Sergipe'
  WHEN seller_state = 'AL' THEN 'Alagoas'
  WHEN seller_state = 'PI' THEN 'Piauí'
  WHEN seller_state = 'AP' THEN 'Amapá'
  WHEN seller_state = 'AM' THEN 'Amazonas'
  WHEN seller_state = 'RR' THEN 'Roraima'
  WHEN seller_state = 'AC' THEN 'Acre'
  WHEN seller_state = 'RO' THEN 'Rondônia'
  WHEN seller_state = 'TO' THEN 'Tocantins'
  WHEN seller_state = 'BA' THEN 'Bahia'
  WHEN seller_state = 'CE' THEN 'Ceará'
  WHEN seller_state = 'DF' THEN 'Distrito Federal'
  WHEN seller_state = 'ES' THEN 'Espírito Santo'
  WHEN seller_state = 'GO' THEN 'Goiás'
  WHEN seller_state = 'MA' THEN 'Maranhão'
  WHEN seller_state = 'MG' THEN 'Minas Gerais'
  WHEN seller_state = 'MS' THEN 'MatoGrosso do Sul'
  WHEN seller_state = 'MT' THEN 'MatoGrosso'
  WHEN seller_state = 'PA' THEN 'Pará'
  WHEN seller_state = 'PB' THEN 'Paraíba'
  WHEN seller_state = 'PE' THEN 'Pernambuco'
  WHEN seller_state = 'PR' THEN 'Paraná'
  WHEN seller_state = 'RJ' THEN 'Rio de Janeiro'
  WHEN seller_state = 'RN' THEN 'Rio Grande do Norte'
  WHEN seller_state = 'RS' THEN 'Rio Grande do Sul'
  WHEN seller_state = 'SC' THEN 'Santa Catarina'
  WHEN seller_state = 'SP' THEN 'São Paulo'
end         AS seller_state,
Round(Avg(Timestamp_diff(order_delivered_customer_date,
```

```
                order_purchase_timestamp
                , day)
            ), 2) AS mean_delivery_time
FROM    `Ecommerce.orders` o
        JOIN `Ecommerce.order_items` oi
          ON o.order_id = oi.order_id
        JOIN `Ecommerce.customers` c
          ON o.customer_id = c.customer_id
        JOIN `Ecommerce.sellers` s
          ON oi.seller_id = s.seller_id
             AND order_status = 'delivered'
GROUP   BY customer_state,
           seller_state
ORDER   BY mean_delivery_time DESC
LIMIT   5;
```

Row	customer_state	seller_state	mean_delivery_time
1	Amazonas	Ceará	138.0
2	Alagoas	Amazonas	90.0
3	Acre	Bahia	66.0
4	Pará	Espírito Santo	36.33
5	Sergipe	Rondônia	36.0

Delivery time is more for orders when buyers and sellers are from different state.

Top 5 States with Lowest Mean Delivery Time:

```
SELECT CASE
          WHEN customer_state = 'SE' THEN 'Sergipe'
          WHEN customer_state = 'AL' THEN 'Alagoas'
          WHEN customer_state = 'PI' THEN 'Piauí'
          WHEN customer_state = 'AP' THEN 'Amapá'
          WHEN customer_state = 'AM' THEN 'Amazonas'
          WHEN customer_state = 'RR' THEN 'Roraima'
          WHEN customer_state = 'AC' THEN 'Acre'
          WHEN customer_state = 'RO' THEN 'Rondônia'
          WHEN customer_state = 'TO' THEN 'Tocantins'
          WHEN customer_state = 'BA' THEN 'Bahia'
          WHEN customer_state = 'CE' THEN 'Ceará'
          WHEN customer_state = 'DF' THEN 'Distrito Federal'
          WHEN customer_state = 'ES' THEN 'Espírito Santo'
          WHEN customer_state = 'GO' THEN 'Goiás'
          WHEN customer_state = 'MA' THEN 'Maranhão'
```

```
        WHEN customer_state = 'MG' THEN 'Minas Gerais'
        WHEN customer_state = 'MS' THEN 'MatoGrosso do Sul'
        WHEN customer_state = 'MT' THEN 'MatoGrosso'
        WHEN customer_state = 'PA' THEN 'Pará'
        WHEN customer_state = 'PB' THEN 'Paraíba'
        WHEN customer_state = 'PE' THEN 'Pernambuco'
        WHEN customer_state = 'PR' THEN 'Paraná'
        WHEN customer_state = 'RJ' THEN 'Rio de Janeiro'
        WHEN customer_state = 'RN' THEN 'Rio Grande do Norte'
        WHEN customer_state = 'RS' THEN 'Rio Grande do Sul'
        WHEN customer_state = 'SC' THEN 'Santa Catarina'
        WHEN customer_state = 'SP' THEN 'São Paulo'
      end         AS customer_state,
      CASE
        WHEN seller_state = 'SE' THEN 'Sergipe'
        WHEN seller_state = 'AL' THEN 'Alagoas'
        WHEN seller_state = 'PI' THEN 'Piauí'
        WHEN seller_state = 'AP' THEN 'Amapá'
        WHEN seller_state = 'AM' THEN 'Amazonas'
        WHEN seller_state = 'RR' THEN 'Roraima'
        WHEN seller_state = 'AC' THEN 'Acre'
        WHEN seller_state = 'RO' THEN 'Rondônia'
        WHEN seller_state = 'TO' THEN 'Tocantins'
        WHEN seller_state = 'BA' THEN 'Bahia'
        WHEN seller_state = 'CE' THEN 'Ceará'
        WHEN seller_state = 'DF' THEN 'Distrito Federal'
        WHEN seller_state = 'ES' THEN 'Espírito Santo'
        WHEN seller_state = 'GO' THEN 'Goiás'
        WHEN seller_state = 'MA' THEN 'Maranhão'
        WHEN seller_state = 'MG' THEN 'Minas Gerais'
        WHEN seller_state = 'MS' THEN 'MatoGrosso do Sul'
        WHEN seller_state = 'MT' THEN 'MatoGrosso'
        WHEN seller_state = 'PA' THEN 'Pará'
        WHEN seller_state = 'PB' THEN 'Paraíba'
        WHEN seller_state = 'PE' THEN 'Pernambuco'
        WHEN seller_state = 'PR' THEN 'Paraná'
        WHEN seller_state = 'RJ' THEN 'Rio de Janeiro'
        WHEN seller_state = 'RN' THEN 'Rio Grande do Norte'
        WHEN seller_state = 'RS' THEN 'Rio Grande do Sul'
        WHEN seller_state = 'SC' THEN 'Santa Catarina'
        WHEN seller_state = 'SP' THEN 'São Paulo'
      end         AS seller_state,
      Round(Avg(Timestamp_diff(order_delivered_customer_date,
                order_purchase_timestamp
                , day)
          ), 2) AS mean_delivery_time
FROM  `Ecommerce.orders` o
      JOIN `Ecommerce.order_items` oi
        ON o.order_id = oi.order_id
      JOIN `Ecommerce.customers` c
        ON o.customer_id = c.customer_id
      JOIN `Ecommerce.sellers` s
```

```
          ON oi.seller_id = s.seller_id
             AND order_status = 'delivered'
GROUP  BY customer_state,
          seller_state
ORDER  BY mean_delivery_time
LIMIT  5;
```

Row	customer_state	seller_state	mean_delivery_time
1	Piauí	Piauí	2.0
2	Paraná	Pará	3.0
3	Minas Gerais	Rio Grande do Norte	3.0
4	Rio Grande do Norte	Rio Grande do Norte	4.38
5	Goiás	Goiás	4.95

Delivery time is less for orders when buyers and sellers are from same state.

- **Top 5 states where delivery is really fast/ not so fast compared to estimated date**

Top 5 States with Delivery is Fast:

```
SELECT CASE
          WHEN customer_state = 'SE' THEN 'Sergipe'
          WHEN customer_state = 'AL' THEN 'Alagoas'
          WHEN customer_state = 'PI' THEN 'Piauí'
          WHEN customer_state = 'AP' THEN 'Amapá'
          WHEN customer_state = 'AM' THEN 'Amazonas'
          WHEN customer_state = 'RR' THEN 'Roraima'
          WHEN customer_state = 'AC' THEN 'Acre'
          WHEN customer_state = 'RO' THEN 'Rondônia'
          WHEN customer_state = 'TO' THEN 'Tocantins'
          WHEN customer_state = 'BA' THEN 'Bahia'
          WHEN customer_state = 'CE' THEN 'Ceará'
          WHEN customer_state = 'DF' THEN 'Distrito Federal'
          WHEN customer_state = 'ES' THEN 'Espírito Santo'
          WHEN customer_state = 'GO' THEN 'Goiás'
          WHEN customer_state = 'MA' THEN 'Maranhão'
          WHEN customer_state = 'MG' THEN 'Minas Gerais'
          WHEN customer_state = 'MS' THEN 'MatoGrosso do Sul'
          WHEN customer_state = 'MT' THEN 'MatoGrosso'
          WHEN customer_state = 'PA' THEN 'Pará'
          WHEN customer_state = 'PB' THEN 'Paraíba'
          WHEN customer_state = 'PE' THEN 'Pernambuco'
          WHEN customer_state = 'PR' THEN 'Paraná'
```

```
        WHEN customer_state = 'RJ' THEN 'Rio de Janeiro'
        WHEN customer_state = 'RN' THEN 'Rio Grande do Norte'
        WHEN customer_state = 'RS' THEN 'Rio Grande do Sul'
        WHEN customer_state = 'SC' THEN 'Santa Catarina'
        WHEN customer_state = 'SP' THEN 'São Paulo'
      end         AS customer_state,
      CASE
        WHEN seller_state = 'SE' THEN 'Sergipe'
        WHEN seller_state = 'AL' THEN 'Alagoas'
        WHEN seller_state = 'PI' THEN 'Piauí'
        WHEN seller_state = 'AP' THEN 'Amapá'
        WHEN seller_state = 'AM' THEN 'Amazonas'
        WHEN seller_state = 'RR' THEN 'Roraima'
        WHEN seller_state = 'AC' THEN 'Acre'
        WHEN seller_state = 'RO' THEN 'Rondônia'
        WHEN seller_state = 'TO' THEN 'Tocantins'
        WHEN seller_state = 'BA' THEN 'Bahia'
        WHEN seller_state = 'CE' THEN 'Ceará'
        WHEN seller_state = 'DF' THEN 'Distrito Federal'
        WHEN seller_state = 'ES' THEN 'Espírito Santo'
        WHEN seller_state = 'GO' THEN 'Goiás'
        WHEN seller_state = 'MA' THEN 'Maranhão'
        WHEN seller_state = 'MG' THEN 'Minas Gerais'
        WHEN seller_state = 'MS' THEN 'MatoGrosso do Sul'
        WHEN seller_state = 'MT' THEN 'MatoGrosso'
        WHEN seller_state = 'PA' THEN 'Pará'
        WHEN seller_state = 'PB' THEN 'Paraíba'
        WHEN seller_state = 'PE' THEN 'Pernambuco'
        WHEN seller_state = 'PR' THEN 'Paraná'
        WHEN seller_state = 'RJ' THEN 'Rio de Janeiro'
        WHEN seller_state = 'RN' THEN 'Rio Grande do Norte'
        WHEN seller_state = 'RS' THEN 'Rio Grande do Sul'
        WHEN seller_state = 'SC' THEN 'Santa Catarina'
        WHEN seller_state = 'SP' THEN 'São Paulo'
      end         AS seller_state,
      Round(Avg(Timestamp_diff(order_delivered_customer_date,
                order_purchase_timestamp
                , day)
          ), 2) AS mean_delivery_time,
      Round(Avg(Timestamp_diff(order_estimated_delivery_date,
                order_purchase_timestamp
                , day)
          ), 2) AS mean_estimated_delivery_time
FROM  `Ecommerce.orders` o
      JOIN `Ecommerce.order_items` oi
        ON o.order_id = oi.order_id
      JOIN `Ecommerce.customers` c
        ON o.customer_id = c.customer_id
      JOIN `Ecommerce.sellers` s
        ON oi.seller_id = s.seller_id
           AND order_status = 'delivered'
GROUP BY customer_state,
```

```
          seller_state
ORDER  BY mean_delivery_time
LIMIT  5;
```

Row	customer_state	seller_state	mean_delivery_time	mean_estimated_delivery_time
1	Piauí	Piauí	2.0	16.0
2	Paraná	Pará	3.0	17.0
3	Minas Gerais	Rio Grande do Norte	3.0	15.0
4	Rio Grande do Norte	Rio Grande do Norte	4.38	19.21
5	Goiás	Goiás	4.95	13.46

States like Piauí, Rio Grande do Norte, and Goiás anticipated delivery times of 15 or more days but only took 3 or 4 days to deliver.

Top 5 States with Delivery is slow:

```
SELECT CASE
         WHEN customer_state = 'SE' THEN 'Sergipe'
         WHEN customer_state = 'AL' THEN 'Alagoas'
         WHEN customer_state = 'PI' THEN 'Piauí'
         WHEN customer_state = 'AP' THEN 'Amapá'
         WHEN customer_state = 'AM' THEN 'Amazonas'
         WHEN customer_state = 'RR' THEN 'Roraima'
         WHEN customer_state = 'AC' THEN 'Acre'
         WHEN customer_state = 'RO' THEN 'Rondônia'
         WHEN customer_state = 'TO' THEN 'Tocantins'
         WHEN customer_state = 'BA' THEN 'Bahia'
         WHEN customer_state = 'CE' THEN 'Ceará'
         WHEN customer_state = 'DF' THEN 'Distrito Federal'
         WHEN customer_state = 'ES' THEN 'Espírito Santo'
         WHEN customer_state = 'GO' THEN 'Goiás'
         WHEN customer_state = 'MA' THEN 'Maranhão'
         WHEN customer_state = 'MG' THEN 'Minas Gerais'
         WHEN customer_state = 'MS' THEN 'MatoGrosso do Sul'
         WHEN customer_state = 'MT' THEN 'MatoGrosso'
         WHEN customer_state = 'PA' THEN 'Pará'
         WHEN customer_state = 'PB' THEN 'Paraíba'
         WHEN customer_state = 'PE' THEN 'Pernambuco'
         WHEN customer_state = 'PR' THEN 'Paraná'
         WHEN customer_state = 'RJ' THEN 'Rio de Janeiro'
         WHEN customer_state = 'RN' THEN 'Rio Grande do Norte'
         WHEN customer_state = 'RS' THEN 'Rio Grande do Sul'
         WHEN customer_state = 'SC' THEN 'Santa Catarina'
         WHEN customer_state = 'SP' THEN 'São Paulo'
       end          AS customer_state,
       CASE
         WHEN seller_state = 'SE' THEN 'Sergipe'
```

```
          WHEN seller_state = 'AL' THEN 'Alagoas'
          WHEN seller_state = 'PI' THEN 'Piauí'
          WHEN seller_state = 'AP' THEN 'Amapá'
          WHEN seller_state = 'AM' THEN 'Amazonas'
          WHEN seller_state = 'RR' THEN 'Roraima'
          WHEN seller_state = 'AC' THEN 'Acre'
          WHEN seller_state = 'RO' THEN 'Rondônia'
          WHEN seller_state = 'TO' THEN 'Tocantins'
          WHEN seller_state = 'BA' THEN 'Bahia'
          WHEN seller_state = 'CE' THEN 'Ceará'
          WHEN seller_state = 'DF' THEN 'Distrito Federal'
          WHEN seller_state = 'ES' THEN 'Espírito Santo'
          WHEN seller_state = 'GO' THEN 'Goiás'
          WHEN seller_state = 'MA' THEN 'Maranhão'
          WHEN seller_state = 'MG' THEN 'Minas Gerais'
          WHEN seller_state = 'MS' THEN 'MatoGrosso do Sul'
          WHEN seller_state = 'MT' THEN 'MatoGrosso'
          WHEN seller_state = 'PA' THEN 'Pará'
          WHEN seller_state = 'PB' THEN 'Paraíba'
          WHEN seller_state = 'PE' THEN 'Pernambuco'
          WHEN seller_state = 'PR' THEN 'Paraná'
          WHEN seller_state = 'RJ' THEN 'Rio de Janeiro'
          WHEN seller_state = 'RN' THEN 'Rio Grande do Norte'
          WHEN seller_state = 'RS' THEN 'Rio Grande do Sul'
          WHEN seller_state = 'SC' THEN 'Santa Catarina'
          WHEN seller_state = 'SP' THEN 'São Paulo'
        end         AS seller_state,
        Round(Avg(Timestamp_diff(order_delivered_customer_date,
                  order_purchase_timestamp
                  , day)
              ), 2) AS mean_delivery_time,
        Round(Avg(Timestamp_diff(order_estimated_delivery_date,
                  order_purchase_timestamp
                  , day)
              ), 2) AS mean_estimated_delivery_time
FROM    `Ecommerce.orders` o
        JOIN `Ecommerce.order_items` oi
          ON o.order_id = oi.order_id
        JOIN `Ecommerce.customers` c
          ON o.customer_id = c.customer_id
        JOIN `Ecommerce.sellers` s
          ON oi.seller_id = s.seller_id
             AND order_status = 'delivered'
GROUP   BY customer_state,
           seller_state
ORDER   BY mean_delivery_time DESC
LIMIT   5;
```

Row	customer_state	seller_state	mean_delivery_time	mean_estimated_delivery_time
1	Amazonas	Ceará	138.0	34.0
2	Alagoas	Amazonas	90.0	52.0
3	Acre	Bahia	66.0	42.0
4	Pará	Espírito Santo	36.33	34.0
5	Sergipe	Rondônia	36.0	41.0

Customer States including Amazonas, Alagoas, and Acre had anticipated delivery days of 34, 52, and 42 days, respectively, however it took 138, 90, and 66 days to deliver.

6. **Payment type analysis:**
    1. **Month over Month count of orders for different payment types**

```
SELECT payment_type,
       Count(CASE
               WHEN month = 1 THEN order_id
             end) AS Jan,
       Count(CASE
               WHEN month = 2 THEN order_id
             end) AS Feb,
       Count(CASE
               WHEN month = 3 THEN order_id
             end) AS Mar,
       Count(CASE
               WHEN month = 4 THEN order_id
             end) AS April,
       Count(CASE
               WHEN month = 5 THEN order_id
             end) AS May,
       Count(CASE
               WHEN month = 6 THEN order_id
             end) AS June,
       Count(CASE
               WHEN month = 7 THEN order_id
             end) AS July,
       Count(CASE
               WHEN month = 8 THEN order_id
             end) AS Aug,
       Count(CASE
               WHEN month = 9 THEN order_id
             end) AS Sep,
       Count(CASE
               WHEN month = 10 THEN order_id
             end) AS Oct,
       Count(CASE
               WHEN month = 11 THEN order_id
             end) AS Nov,
```

```
        Count(CASE
                WHEN month = 12 THEN order_id
              end) AS DEC
FROM   (SELECT payment_type,
               p.order_id,
               Extract(month FROM order_purchase_timestamp) AS month
        FROM   `Ecommerce.payments` p
               JOIN `Ecommerce.orders` o
                 ON p.order_id = o.order_id
        WHERE  order_status = 'delivered')x
GROUP  BY payment_type
ORDER  BY Count(order_id) DESC
```

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS

Row	payment_type	Jan	Feb	Mar	April	May	June	July	Aug	Sep	Oct	Nov	Dec
1	credit_card	5910	6371	7434	7113	8131	7133	7634	8090	3183	3625	5716	4246
2	UPI	1661	1665	1881	1739	1982	1778	2011	2021	868	1006	1445	1134
3	voucher	461	408	578	532	598	556	618	514	277	296	367	288
4	debit_card	118	81	104	119	78	206	254	303	43	53	65	62

Customers prefer paying via credit card and UPI over other options that might be due to discounted offers of credit card and ease accessibility of UPI.

Sales Percentage v/s Payment Methods

```
SELECT   payment_type,
         Count(order_id)
          AS TotalOrders,
         Round(( Count(order_id) * 100 ) / ( Sum(Count(order_id)) OVER
 () ), 2) AS percentage_of_orders
FROM     (
                SELECT payment_type,
                       p.order_id,
                FROM   `Ecommerce.payments` p
                JOIN   `Ecommerce.orders` o
                ON     p.order_id = o.order_id
                WHERE  order_status = 'delivered')x
GROUP BY payment_type
ORDER BY TotalOrders DESC
```

Row	payment_type	TotalOrders	percentage_of_orders
1	credit_card	74586	74.03
2	UPI	19191	19.05
3	voucher	5493	5.45
4	debit_card	1486	1.47

94% Payment were done via Credit card (74%) and UPI (19%).

2. Distribution of payment installments and count of orders

```
SELECT payment_installments,
       Count(order_id) AS count_of_orders
FROM   `Ecommerce.payments`
GROUP  BY payment_installments
ORDER  BY Count(order_id) DESC
```

JOB INFORMATION | RESULTS | JSON

Row	payment_installments	count_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

Most clients choose fewer instalments when paying with a credit card to avoid paying interest.

- **Analysis on Products and Rating**

Top 10 frequently ordered item:

```
SELECT product_category,
       Count(o.order_id)            AS Ordercount,
       Round(Avg(review_score), 1) AS avg_rating
FROM   `Ecommerce.order_items` oi
       JOIN `Ecommerce.products` p
         ON oi.product_id = p.product_id
       JOIN `Ecommerce.orders` o
         ON oi.order_id = o.order_id
       JOIN `Ecommerce.order_reviews` r
         ON o.order_id = r.order_id
WHERE  order_status = 'delivered'
GROUP  BY product_category
ORDER  BY ordercount DESC;
```

JOB INFORMATION | RESULTS | JSON | EXECUTION DET

Row	product_category	Ordercount	avg_rating
1	bed table bath	10985	3.9
2	HEALTH BEAUTY	9456	4.2
3	sport leisure	8436	4.2
4	Furniture Decoration	8159	4.0
5	computer accessories	7672	4.0
6	housewares	6780	4.1
7	Watches present	5825	4.1
8	telephony	4408	4.0
9	Garden tools	4254	4.1
10	automotive	4117	4.1

Product categories include bed table bath, health beauty, and sports leisure had higher sales and an average rating of 4.

Top 10 ordered item with Highest Rating:

```
SELECT product_category,
       Count(o.order_id)            AS Ordercount,
       Round(Avg(review_score), 1) AS avg_rating
FROM   `Ecommerce.order_items` oi
       JOIN `Ecommerce.products` p
         ON oi.product_id = p.product_id
       JOIN `Ecommerce.orders` o
         ON oi.order_id = o.order_id
       JOIN `Ecommerce.order_reviews` r
         ON o.order_id = r.order_id
WHERE  order_status = 'delivered'
GROUP  BY product_category
ORDER  BY avg_rating DESC;
```

Row	product_category	Ordercount	avg_rating
1	Fashion Children's Clothing	7	5.0
2	cds music dvds	14	4.6
3	General Interest Books	533	4.5
4	Imported books	57	4.5
5	Drink foods	271	4.4
6	Construction Tools Tools	99	4.4
7	Bags Accessories	1073	4.4
8	technical books	264	4.4
9	flowers	31	4.4
10	HOUSE PASTALS OVEN AND C...	73	4.4

Few products, such as children's apparel, music CDs and DVDs, and general interest publications, have excellent ratings but poor sales.

Distribution of products across Top 5 states

```
SELECT    product_category,
          Sum(
          CASE
                    WHEN customer_state='SP' THEN 1
                    ELSE 0
          END) AS SP,
          Sum(
          CASE
                    WHEN customer_state='RJ' THEN 1
                    ELSE 0
          END) AS RJ,
          Sum(
          CASE
                    WHEN customer_state='MG' THEN 1
                    ELSE 0
          END) AS MG,
          Sum(
          CASE
                    WHEN customer_state='RS' THEN 1
                    ELSE 0
          END) AS RS,
          Sum(
          CASE
                    WHEN customer_state='PR' THEN 1
                    ELSE 0
          END) AS PR,
FROM      `Ecommerce.order_items` oi
JOIN      `Ecommerce.products` p
ON        oi.product_id=p.product_id
JOIN      `Ecommerce.orders` o
ON        oi.order_id=o.order_id
JOIN      `Ecommerce.customers` c
ON        o.customer_id=c.customer_id
GROUP BY product_category
ORDER BY SP DESC;
```

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS

Row	product_category	SP	RJ	MG	RS	PR
1	bed table bath	5235	1644	1331	614	468
2	HEALTH BEAUTY	4204	1064	1086	436	425
3	sport leisure	3667	1041	966	475	486
4	Furniture Decoration	3531	1090	949	561	520
5	housewares	3265	877	835	434	337
6	computer accessories	3170	1002	1000	485	389
7	Watches present	2281	874	637	233	307
8	automotive	1747	442	513	188	234
9	toys	1712	585	494	212	205
10	telephony	1646	438	484	292	250

Popular states like São Paulo, Rio de Janeiro, and Minas Gerais receive more orders for product categories that have greater sales, like bed table bath, health beauty, and sports leisure.

Distribution of most selling products across Top cities

```
SELECT product_category,
       Sum(CASE
             WHEN customer_city = 'sao paulo' THEN 1
             ELSE 0
           end) AS sao_paulo,
       Sum(CASE
             WHEN customer_city = 'rio de janeiro' THEN 1
             ELSE 0
           end) AS rio_de_janeiro,
       Sum(CASE
             WHEN customer_city = 'belo horizonte' THEN 1
             ELSE 0
           end) AS belo_horizonte,
       Sum(CASE
             WHEN customer_city = 'brasilia' THEN 1
             ELSE 0
           end) AS brasilia,
       Sum(CASE
             WHEN customer_city = 'curitiba' THEN 1
             ELSE 0
           end) AS curitiba
```

```
FROM    `Ecommerce.order_items` oi
        JOIN `Ecommerce.products` p
          ON oi.product_id = p.product_id
        JOIN `Ecommerce.orders` o
          ON oi.order_id = o.order_id
        JOIN `Ecommerce.customers` c
          ON o.customer_id = c.customer_id
GROUP   BY product_category
ORDER   BY sao_paulo DESC;
```

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS

Row	product_category	sao_paulo	rio_de_janeiro	belo_horizonte	brasilia	curitiba
1	bed table bath	1984	854	351	206	144
2	HEALTH BEAUTY	1753	572	274	246	138
3	sport leisure	1397	593	214	222	152
4	housewares	1321	488	245	148	117
5	Furniture Decoration	1268	615	246	152	151
6	computer accessori...	1227	538	256	163	145
7	Watches present	844	470	129	148	110
8	toys	658	350	143	97	83
9	telephony	620	216	97	94	67
10	automotive	555	195	105	109	55

Insights:

1) All the unique Id's, cities and states are of type String, Ordered, Delivered and other Dates are of type Timestamp whereas attributes like price, payment, freight value ae of type float.
2) The Dataset includes 27 states and a total of 8011 distinct cities of Brazil.
3) When there are big events like Black Friday, Thanksgiving, Christmas, and New Year's and sales were at their zenith during these periods, an increasing tendency is noted in e-commerce. There have also been a few months where sales were weak but spiked during sale periods like Father's Day and Women's Day.
4) Customers purchase more than 75% of all orders at night and in the afternoon.
5) States from the southeast, such São Paulo, Rio de Janeiro, and Minas, have the highest sales rates compared to other states and regions.

6) When customers place orders from states like São Paulo, Paraná, Minas Gerais, and Rio de Janeiro or if sellers are from states like São Paulo, Rio Grande do Sul, Rio Grande do Norte, and Mato Grosso do Sul, delivery times are reduced.
7) Customers prefer using credit cards and UPI over other methods of payment, maybe because of the credit cards' discounted offers and UPI's simplicity of use.
8) Popular states like São Paulo, Rio de Janeiro, and Minas Gerais and their capitals receive more orders for product categories that have greater sales, like bed table bath, health beauty, and sports leisure.
9) The states of the buyer and seller affect the freight value and delivery time. The delivery time and cost for transportation, loading, and unloading of goods decrease when these states are closer together/ or same and increase if they are at distant or different.

Recommendations:

1) Order payments comprised 74% credit card payments and 19% UPI payments. Customers prefer to pay with credit cards and UPI, thus more promotions can be introduced to these payment options to attract more clients.
2) Due to the holiday seasons, sales are higher in the beginning and end of the year but drop in the middle. A mid-year event or some amazing mid-year deal can be introduced to boost the sales during this period.
3) Customers wait a long time to place orders on Black Friday and Thanksgiving Day, and 37% of transactions were made during the sales week, yet the average delivery time is 15+ days during that time. To draw in more customers during this time, delivery times can be shortened.
4) Around 75% of sales occur throughout the night and afternoon, which is the peak period for sales, thus if any sales are being introduced, these times should be chosen.
5) There are many orders that were meant to be delivered in 20 to 30 days but ended up taking more than 100 days. These deliveries can be made better, especially in states like Amapá, Roraima, Amazonas, and Alagoas where the average delivery time is longer than in other states.
6) Children's apparel, music CDs and DVDs, general interest periodicals, food, and beverages are examples of products with high product ratings but low sales. These products might be advertised in order to boost sales. While bedside tables, bath accessories, health and beauty products, and sports and leisure items sell well, their ratings might be raised by focusing on product quality.