1. Data understanding
2. orders

Whether there are repetitive order\_id per each customer\_id?

SELECT

COUNT(order\_id) AS number\_order,

COUNT(customer\_id) AS number\_customer

FROM orders;

(99441, 99441)

* count the number of orders per status

SELECT

COUNT(order\_id) AS number\_delivered

FROM orders

WHERE order\_status = 'delivered';

(96478)

SELECT

COUNT(order\_id) AS number\_delivered

FROM orders

WHERE order\_status = 'shipped';

(1107)

SELECT

COUNT(order\_id) AS number\_delivered

FROM orders

WHERE order\_status = 'canceled';

(625)

SELECT

COUNT(order\_id) AS number\_delivered

FROM orders

WHERE order\_status = 'unavailable';

(609)

SELECT

COUNT(order\_id) AS number\_delivered

FROM orders

WHERE order\_status = 'invoiced';

(314)

What the specific order status mean:

* delivered: delivery to the customer finished
* shipped: payment finished and delivery started accordingly.
* invoiced / approved / processing: payment finished and waiting for the start of delivery
* unavailable: It might be the cancellation from the seller's side
* canceled: the canceled order would be the cancellation from the customer's side.

1. Payments

SELECT

COUNT(distinct order\_id) AS distinct\_orders,

COUNT(payment\_sequential) AS total\_payment,

COUNT(payment\_sequential) - COUNT(distinct order\_id) AS duplicated\_rows

FROM payments;

(99440, 103886, 4446)

There are some customers pay an order with more than one payment method.

There are 5 methods for payment\_type:

SELECT DISTINCT payment\_type

FROM payments;

(credit\_card, boleto, voucher, debit\_card, not\_defined)

1. Reviews

There are more than 800 duplicate rows on review\_id. If review\_id is same => every content except “order\_id” is the same.

SELECT

COUNT(DISTINCT review\_id),

COUNT(review\_id)

FROM reviews;

(98410, 99224)

There are more than 500 duplicate rows on order\_id in reviews table. If order\_id is same => every content is different from each other.

SELECT

COUNT(DISTINCT order\_id),

COUNT(order\_id)

FROM reviews;

(98673, 99224)

There are null values on 'review\_comment\_title' and 'review\_comment\_message': 88.3%, 58.7% are null values respectively.

SELECT COUNT(\*) AS total\_rows,

(SELECT COUNT(\*) FROM reviews WHERE review\_comment\_title IS NULL) AS total\_null

FROM reviews;

(99224, 87656)

SELECT COUNT(\*) AS total\_rows,

(SELECT COUNT(\*) FROM reviews WHERE review\_comment\_message IS NULL) AS total\_null

FROM reviews;

(99224, 58247)

1. Products & english\_product\_name

There aren’t duplicate rows on product\_id. Total 73 category names, 32,951 products.

SELECT COUNT(\*),

COUNT(DISTINCT product\_id),

COUNT(distinct product\_category\_name)

FROM products;

(32951, 32951, 73)

There are some products that are not in any category (NULL value at product\_category\_name).

SELECT product\_id,

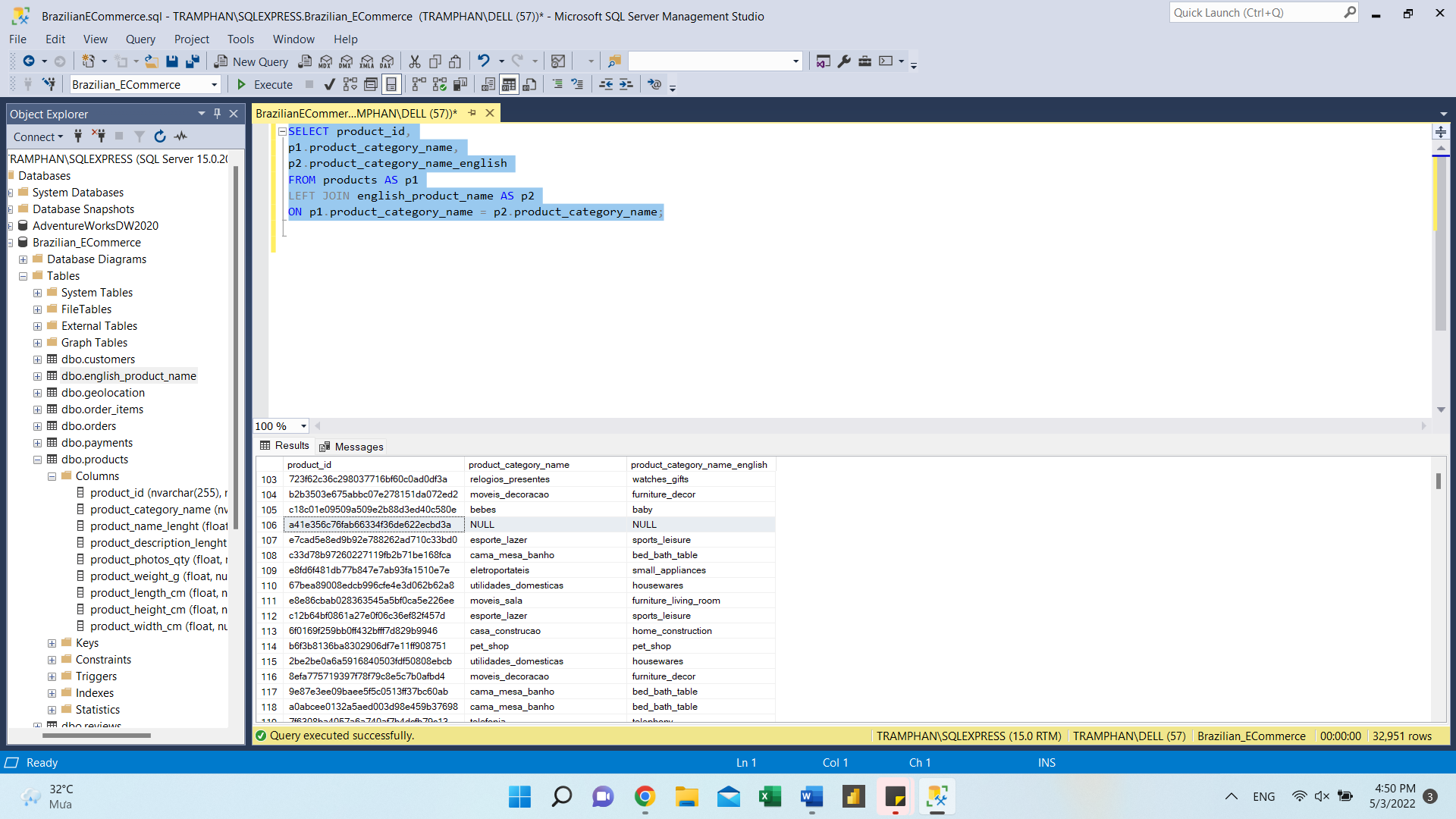
p1.product\_category\_name,

p2.product\_category\_name\_english

FROM products AS p1

LEFT JOIN english\_product\_name AS p2

ON p1.product\_category\_name = p2.product\_category\_name;



1. Sellers

There aren't duplicated rows on 'seller\_id'. Total 3095 sellers.

SELECT COUNT(\*),

COUNT(DISTINCT seller\_id)

FROM sellers;

(3095, 3095)

1. Customers

There aren't duplicated rows on 'customer\_id'. Total 99441 sellers.

SELECT COUNT(\*),

COUNT(DISTINCT customer\_id)

FROM customers;

(99441, 99441)

The distinct number of customers: 96096.

SELECT

COUNT(DISTINCT customer\_unique\_id)

FROM customers;

(96096)