E-Commerce Business Performance Analysis | Data Analysis |

Business performance analysis with SQL.

Source: Freepik

About the project:

Any business must understand its business performance and develop business improvement ideas and activities. In this project, I’m going to analyze the business

performance of an **E-commerce** company and try to find insights into its customer growth, product sales, and payment methods.

It has information on 100k orders from 2016 to 2018 made at multiple marketplaces in Brazil. Its features allow viewing orders from various dimensions: from order status, price, payment, and freight performance to customer location, product attributes, and reviews written by customers. I will be performing the analysis using **Microsoft SQL Server** and **Power BI**.

Objectives:

With the data that is provided, we would like to know about

i) Overall customer activity growth from 2016 to 2018 by seeing average active users, new customers, customers with repeat orders, and average orders by customers.

ii) Overall product category quality from 2016 to 2018 by seeing total revenue, total canceled orders, best-selling product category, and most canceled product category.

iii) Overall payment type usage from 2016 to 2018 by seeing favorite payment type all time, and the amount of usage for each type of payment by year.

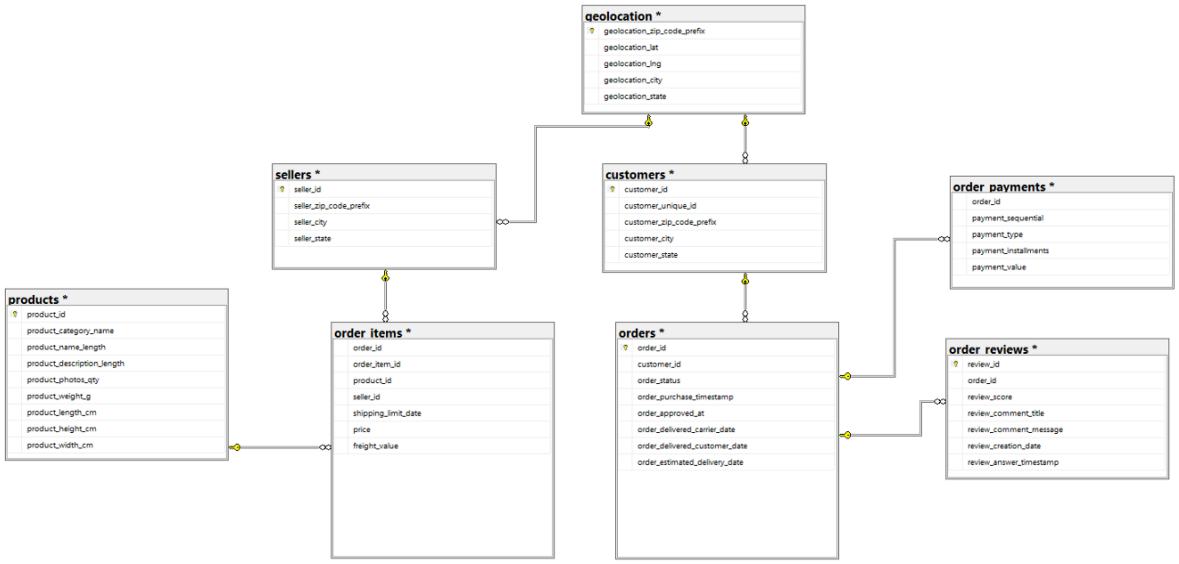
About the Dataset:

There are 8 datasets included in this project for the analysis namely:

1. customers
2. geolocation
3. order_items
4. order_payments
5. order_reviews
6. orders
7. products
8. sellers

You can access the dataset **here.**

Relationship Diagram:



Entity Relationship Diagram using SQL Server

Analysis:

The whole SQL query is available on my ***GitHub profile***.

I) Overall customer activity growth from 2016 to 2018

1. **Average active users per year**

```
SELECT
  year,
  ROUND(AVG(total_customers), 0) avg_active_users
FROM
 (
   SELECT
     DATEPART(YEAR, o.order_purchase_timestamp) year,
     DATEPART(MONTH, o.order_purchase_timestamp) month,
     COUNT(DISTINCT(c.customer_unique_id)) total_customers
   FROM
     orders o
      JOIN customers c
        ON o.customer_id = c.customer_id
   GROUP BY
     DATEPART(YEAR, o.order_purchase_timestamp),
```

```
    DATEPART(MONTH, o.order_purchase_timestamp)
 ) total_users
GROUP BY
  year;
```

Output:

	year	avg_active_users
1	2016	108
2	2017	3694
3	2018	5338

2. New customers per year

```
SELECT
  DATEPART(YEAR, first_order) year,
  COUNT(DISTINCT(customer_unique_id)) new_customers
FROM
 (
   SELECT
     c.customer_unique_id,
     MIN(o.order_purchase_timestamp) first_order
   FROM
     customers c
      JOIN orders o
        ON c.customer_id = o.customer_id
   GROUP BY
     c.customer_unique_id
 ) first_purchase
GROUP BY
  DATEPART(YEAR, first_order);
```

	year	new_customers
1	2016	326
2	2017	43708
3	2018	52062

3. No.of customers with repeat orders

```
SELECT
  year,
  COUNT(customer_unique_id) customers_with_repeat_orders
FROM
 (
   SELECT
     DATEPART(YEAR, o.order_purchase_timestamp) year,
     c.customer_unique_id,
     COUNT(o.order_id) no_of_orders
   FROM
     customers c
      JOIN orders o
        ON c.customer_id = o.customer_id
   GROUP BY
     DATEPART(YEAR, o.order_purchase_timestamp),
     c.customer_unique_id
   HAVING
     COUNT(o.order_id) > 1
 ) ro
GROUP BY
  year;
```

	year	customers_with_repeat_orders
1	2016	3
2	2017	1256
3	2018	1167

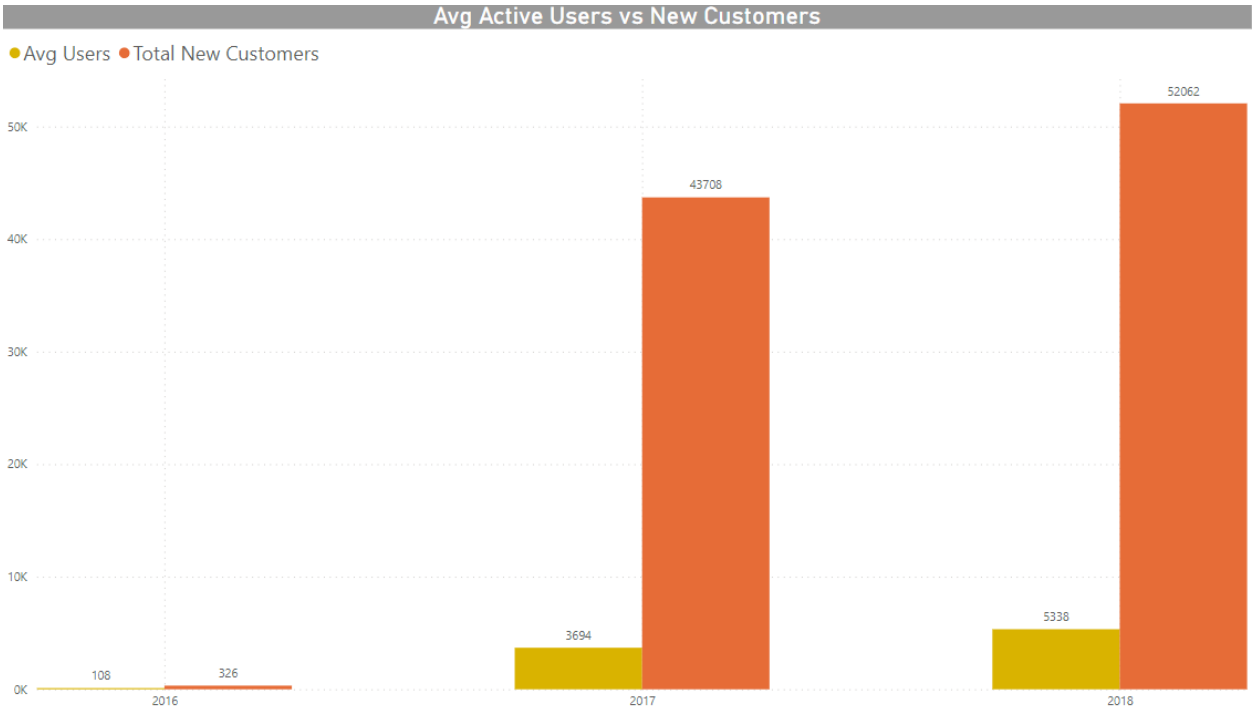
4. Average no.of orders by customers

```
SELECT
  year,
  AVG(no_of_orders) avg_orders_by_customer
FROM
 (
   SELECT
     DATEPART(YEAR, o.order_purchase_timestamp) year,
     c.customer_unique_id,
     COUNT(o.order_id) no_of_orders
   FROM
     customers c
      JOIN orders o
        ON c.customer_id = o.customer_id
```

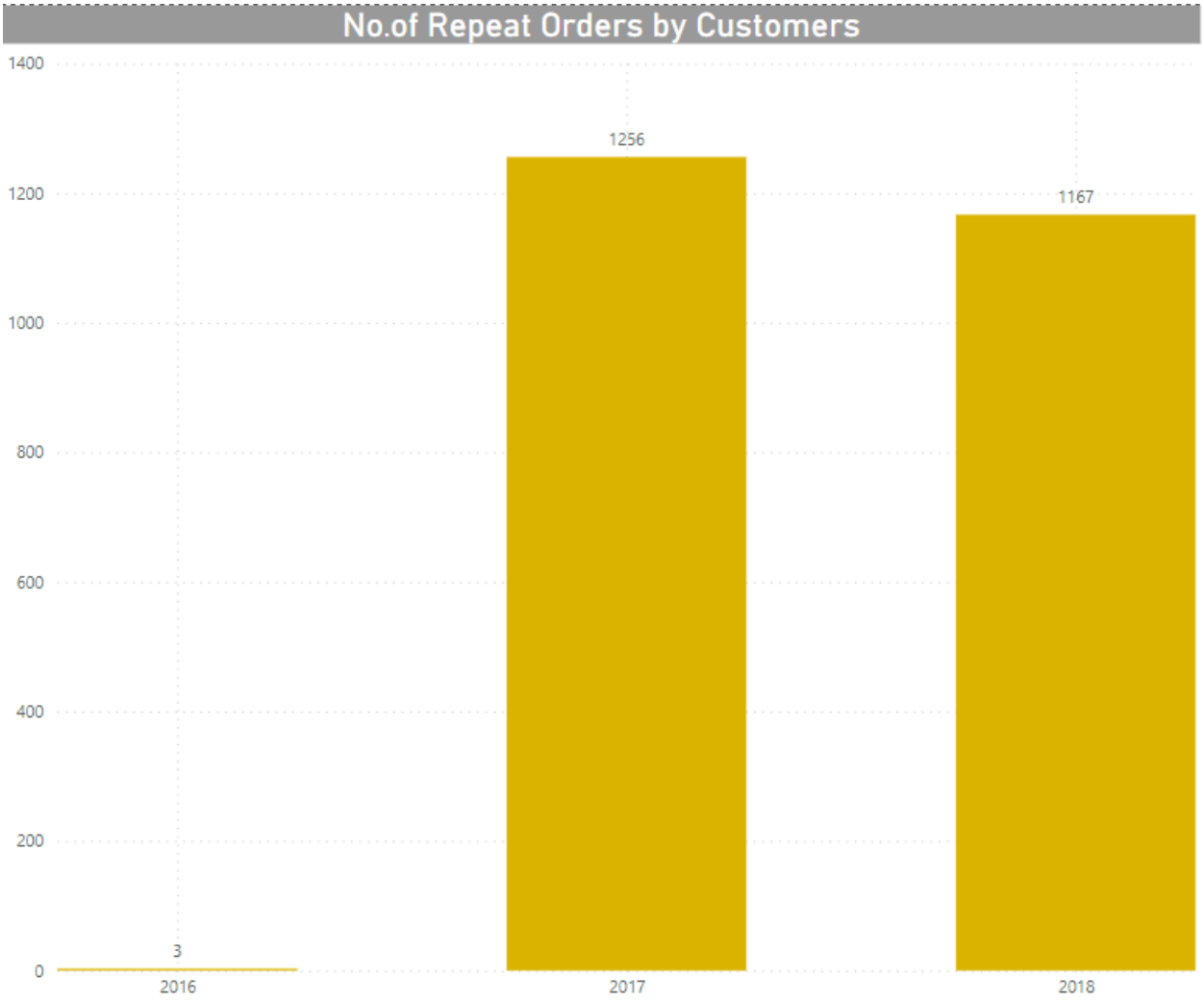
```
    GROUP BY
      DATEPART(YEAR, o.order_purchase_timestamp),
      c.customer_unique_id
  ) repeat_orders
GROUP BY
  year;
```

	year	avg_orders_by_customer
1	2016	1
2	2017	1
3	2018	1

Avg no.of orders by customers



- *From the above diagram, we can see that there is a significant increase in the number of Avg Active Users as well as New Customers from 2016 to 2018.*
- *These numbers are comparatively smaller in 2016. This might be because the datasets may not contain enough information for 2016.*



- *2017 seems to be the year with the most no.of repeat orders.*
- *There was a significant increase in the no.of repeat orders from 2016 to 2017 but the numbers seem to be lower in 2018 than that in 2017.*
- *There is only a slight difference in the numbers in 2017 and 2018 because the average order per customer is only 1.*

II)Product category quality analysis 2016–2018

1. **Total revenue per year**

```
SELECT
  DATEPART(YEAR, o.order_purchase_timestamp) year,
  ROUND(SUM(rev.revenue_per_order), 2) revenue
FROM
 (
   SELECT
     order_id,
     ROUND(SUM(price + freight_value), 2) revenue_per_order
   FROM
     order_items
   GROUP BY
     order_id ) rev
  JOIN orders o
   ON rev.order_id = o.order_id
WHERE
  o.order_status = 'delivered'
GROUP BY
  DATEPART(YEAR, o.order_purchase_timestamp)
ORDER BY
  DATEPART(YEAR, o.order_purchase_timestamp) DESC;
```

	year	revenue
1	2018	8451584.77
2	2017	6921535.24
3	2016	46653.74

2. Total canceled orders per year

```
SELECT
  DATEPART(YEAR, order_purchase_timestamp) year,
  COUNT(DISTINCT order_id) canceled_orders
FROM
  orders
WHERE
  order_status = 'canceled'
GROUP BY
  DATEPART(YEAR, order_purchase_timestamp)
ORDER BY
  DATEPART(YEAR, order_purchase_timestamp) DESC
```

	year	canceled_orders
1	2018	334
2	2017	265
3	2016	26

3. Best selling product category by year

```
SELECT
  year,
  product_category,
  revenue
FROM
 (
   SELECT
     DATEPART(YEAR, o.order_purchase_timestamp) year,
     p.product_category_name product_category,
     ROUND(SUM(oi.price + oi.freight_value), 2) revenue,
     RANK() OVER
           (
            PARTITION BY DATEPART(YEAR, o.order_purchase_timestamp)
            ORDER BY ROUND(SUM(oi.price + oi.freight_value), 2) DESC
           ) AS ranking
FROM
  products p
   JOIN order_items oi
     ON p.product_id = oi.product_id
   JOIN orders o
     ON oi.order_id = o.order_id
GROUP BY
  DATEPART(YEAR, o.order_purchase_timestamp),
  p.product_category_name
 ) revenue_rank
WHERE
  ranking = 1
ORDER BY
  year DESC
```

	year	product_category	revenue
1	2018	Health Beauty	885191.12
2	2017	Bed Bath Table	590280.44
3	2016	Furniture Decor	7188.51

4. Most canceled category

```
SELECT
  year,
  product_category,
  canceled_orders
FROM
 (
   SELECT
     DATEPART(YEAR, o.order_purchase_timestamp) year,
     p.product_category_name product_category,
     COUNT(o.order_id) canceled_orders,
     RANK() OVER
            (
             PARTITION BY DATEPART(YEAR, o.order_purchase_timestamp)
             ORDER BY COUNT(o.order_id) DESC
            ) ranking
  FROM
    orders o
     JOIN order_items oi
       ON o.order_id = oi.order_id
     JOIN products p
       ON oi.product_id = p.product_id
  WHERE
    o.order_status = 'canceled'
  GROUP BY
    DATEPART(YEAR, order_purchase_timestamp),
    p.product_category_name
 ) cancels
WHERE
  ranking = 1
ORDER BY
  year DESC
```

	year	product_category	canceled_orders
1	2018	Health Beauty	27
2	2017	Sports Leisure	25
3	2016	Toys	3

Year	Revenue
2018	$84,51,584.77
2017	$69,21,535.24
2016	$46,653.74
Total	**$1,54,19,773.75**

- *From **2016 to 2018** there is a significant **increase in revenue**.*
- ***2018** is the year with the **highest revenue.***

Best Selling Category & Revenue by Year	
Furniture Decor	
2016	$7,188.51
Bed Bath Table	
2017	$5,90,280.44
Health Beauty	
2018	$8,85,191.12

Most Cancelled Cateory		
Year	Category	Cancelled Orders
2018	Health Beauty	334
2017	Sports Leisure	265
2016	Toys	26
Total		**625**

- **Best selling categories** based on revenue in **2016**, **2017**, and **2018** are “**Furniture Decor**”, “**Bed Bath Table**”, and “**Health & Beauty**” respectively.
- Also, the “**Health & Beauty**” category is the **best selling** as well as the category with the **most cancellations** in **2018**.

III) Payment Type Usage Analysis

1. **Favorite payment type**

```
SELECT
  payment_type,
  COUNT(payment_type) no_of_usage
FROM
  order_payments
GROUP BY
  payment_type
ORDER BY
  no_of_usage DESC
```

	payment_type	no_of_usage
1	credit_card	76795
2	boleto	19784
3	voucher	5775
4	debit_card	1529
5	not_defined	3

2. Top favorite payment type by year

```
SELECT
  year,
  favorite_payment_type,
  no_of_usage
FROM
 (
   SELECT
     DATEPART(YEAR, o.order_purchase_timestamp) year,
     op.payment_type favorite_payment_type,
     COUNT(op.payment_type) no_of_usage,
     RANK() OVER
            (
             PARTITION BY DATEPART(YEAR, o.order_purchase_timestamp)
             ORDER BY COUNT(op.payment_type) DESC
            ) payment_ranking
   FROM
     order_payments op
      JOIN orders o
        ON op.order_id = o.order_id
   GROUP BY
     DATEPART(YEAR, o.order_purchase_timestamp),
     op.payment_type
 ) p
WHERE
  payment_ranking = 1
```

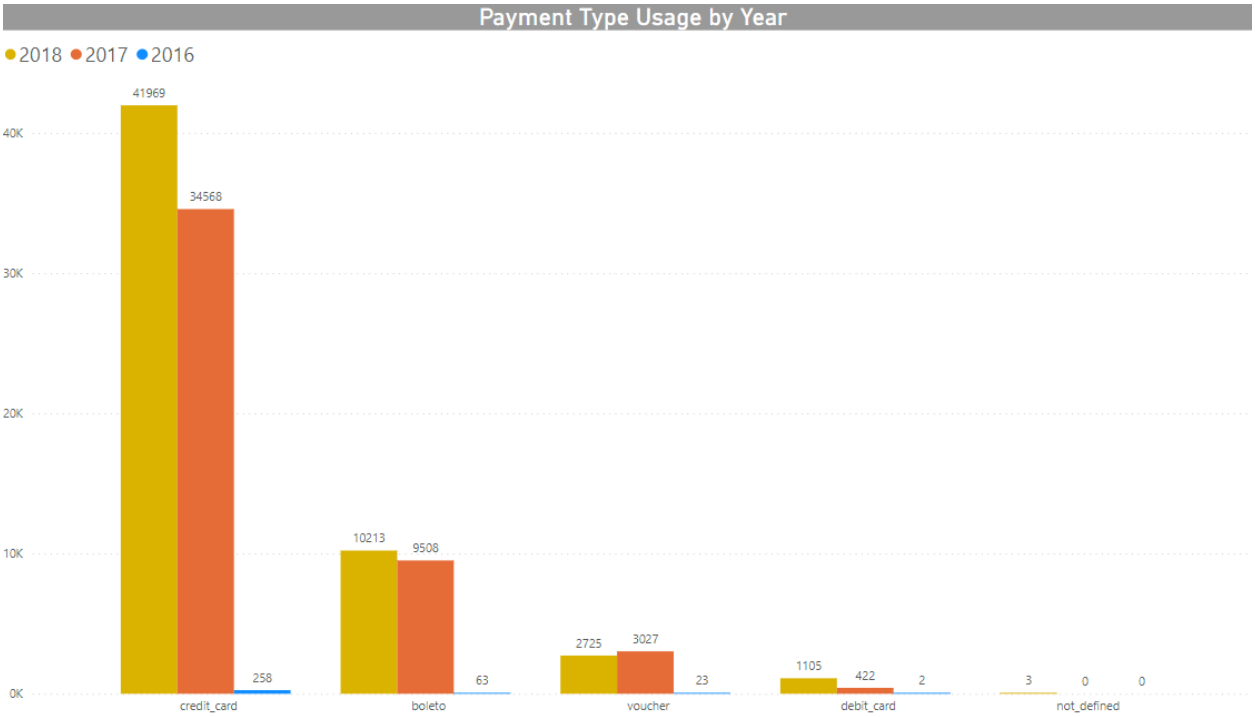
	year	favorite_payment_type	no_of_usage
1	2016	credit_card	258
2	2017	credit_card	34568
3	2018	credit_card	41969

3. Payment usage per year

```
WITH usage AS
 (
   SELECT
     DATEPART(YEAR, o.order_purchase_timestamp) year,
     op.payment_type payment_type,
```

```
    COUNT(op.payment_type) no_of_usage
  FROM
    order_payments op
     JOIN orders o
       ON op.order_id = o.order_id
  GROUP BY
    DATEPART(YEAR, o.order_purchase_timestamp),
    op.payment_type
 )SELECT
  payment_type,
SUM(CASE WHEN year = 2016 THEN no_of_usage ELSE 0 END) '2016_usage',
SUM(CASE WHEN year = 2017 THEN no_of_usage ELSE 0 END) '2017_usage',
SUM(CASE WHEN year = 2018 THEN no_of_usage ELSE 0 END)'2018_usage'
FROM
  usage
GROUP BY
  payment_type
```

	payment_type	2016_usage	2017_usage	2018_usage
1	boleto	63	9508	10213
2	credit_card	258	34568	41969
3	debit_card	2	422	1105
4	not_defined	0	0	3
5	voucher	23	3027	2725



- ***Credit card*** *is the* ***most preferred method of payment*** *throughout the years.*
- ***Credit cards*** *accounted for* ***74.92%*** *of the transactions in* ***2018.***
- *Debit cards also saw a significant increase in usage from 2017 to 2018.*
- *Vouchers have a slight decline in usage in 2018 compared to 2017.*

Conclusion:

From the analysis, we can conclude that:

- In terms of Annual Customer Growth, there is an increase in the number of Average Active Users and New Customers from 2016 to 2018.
- On the other hand, 2018 saw a slight decline in the number of customers who purchased more than one product compared to 2017.
- 2018 is the year with the highest revenue(54.81%) followed by 2017(44.89%) and 2016(0.30%). The revenue has been increasing since 2016.
- “Health & Beauty” is the best-selling category in 2018 with a revenue of $ 8,85,191.12 followed by “Bed Bath Table” in 2017 ($ 5,90,280.44) and “Furniture Decor” in 2016 ($ 7,188.51).

- “Health & Beauty” is also the category with the most cancellations in 2018 followed by “Sports Leisure” in 2017 and “Toys” in 2016.
- Credit cards dominate the method of payment throughout the years: 74.92% in 2018, 72.74% in 2017, and 74.57% in 2016.