Analyzing eCommerce Business Performance with SQL

Created by:
Muhammad Rizki Mardanu Hilman
Your Email : rizkimardanu9@gmail.com
linkedIn : linkedin.com/in/rizkimardanu

Stay curious, keep exploring. Graduated from Bandung Institute of Technology. Passionate about Data Science and data analysis. Problem finding and process-oriented data analyst with in-depth knowledge of machine learning, big data capture, analyzing and processing data using BI tools.

Had experience handling big data using SQL, Google Data Studio, and Python on project programming as a Bootcamp Candidate. lovely experience executing some BI tools; jupyter notebook, spyder to Processing Data, Statistics, Data Visualization, and Machine Learning. This moment helped me gain confidence and faith in data scientists with result-based thinking is a must within an organization. Remember something that you can't measure, you can't control. In short, a well-performed Data Scientist with result-based thinking, strong attention to detail, good execution using tools, and an initiative person will help to provide indicators and information to strengthen, sustainables and organize your organization.

- In a company measuring business performance is very important to track, monitor, and assess the success or failure of various business processes. This can help us to see current market conditions, growth analysis, and product analysis, and to develop opportunities for new, more effective business methods. Therefore, this project will analyze the business performance of an eCommerce company, considering several business metrics, namely customer growth, product quality, and payment types.

- In this project, an analysis will be carried out using **PostgreSQL**,

Rakamin Academy

1. Creating Database & Table :

● Create Database

```
CREATE DATABASE ecommerce;
```

● Create Tables

Create several tables in the ecommerce database along with column names, data types, and primary keys.

```
CREATE TABLE products (
    column1 int4 NULL,
    product_id varchar(50) NOT NULL,
    product_category_name varchar(50) NULL,
    product_name_lenght float8 NULL,
    product_description_lenght float8 NULL,
    product_photos_qty float8 NULL,
    product_weight_g float8 NULL,
    product_length_cm float8 NULL,
    product_height_cm float8 NULL,
    product_width_cm float8 NULL,
    CONSTRAINT products_pk PRIMARY KEY (product_id)
);
```

List of Tables Created
customers
order_reviews
geolocation
orders
order_items
products
order_payments
sellers

2. Importing Data :

● Import Data into Table

1. Klik kanan pada tabel yang ingin diisi.
2. Pilih Import/Export
3. Pilih Import
4. Pilih file CSV yang sudah ada di komputer Anda
5. Pilih Header = Yes jika baris pertama CSV merupakan nama kolom
6. Pilih delimiter yang sesuai dengan file CSV
7. Klik OK

● Pre-Cleaning Data

Specifically for geolocation data, cleaning will be carried out first
including :

1. Drop Duplicate Rows
2. Change Special Character in City
3. Input new geolocations from customers and sellers

```
-- create geolocation clean
CREATE TABLE geolocation_dirty2 AS
SELECT geolocation_zip_code_prefix, geolocation_lat, geolocation_lng,
REPLACE(REPLACE(REPLACE(
TRANSLATE(TRANSLATE(TRANSLATE(TRANSLATE(
TRANSLATE(TRANSLATE(TRANSLATE(TRANSLATE(
    geolocation_city, '£,³,´,.', ''), '`', ''''),
    'é,ê', 'e,e'), 'á,â,ã', 'a,a,a'), 'ô,ó,õ', 'o,o,o'),
    'ç', 'c'), 'ú,ü', 'u,u'), 'í', 'i'),
    '4o', '4º'), '* ', ''), '%26apos%3b', ''''
) AS geolocation_city, geolocation_state
from geolocation_dirty gd;

CREATE TABLE geolocation AS
WITH geolocation AS (
    SELECT geolocation_zip_code_prefix,
    geolocation_lat,
    geolocation_lng,
    geolocation_city,
    geolocation_state FROM (
        SELECT *,
            ROW_NUMBER() OVER (                custgeo AS (
                PARTITION BY geoloca         SELECT customer_zip_code_prefix, geolocation_lat,
            ) AS ROW_NUMBER              geolocation_lng, customer_city, customer_state
        FROM geolocation_dirty2          FROM (
    ) TEMP                                   SELECT *,
    WHERE ROW_NUMBER = 1                         ROW_NUMBER() OVER (
),                                                   PARTITION BY customer_zip_code_prefix
                                                 ) AS ROW_NUMBER
                                             FROM (
                                                 SELECT customer_zip_code_prefix, geolocation_lat,
                                                 geolocation_lng, customer_city, customer_state
                                                 FROM customers cd
                                                 LEFT JOIN geolocation_dirty gdd
                                                 ON customer_city = geolocation_city
                                                 AND customer_state = geolocation_state
                                                 WHERE customer_zip_code_prefix NOT IN (
                                                     SELECT geolocation_zip_code_prefix
                                                     FROM geolocation gd
                                                 )
                                             ) geo
                                         ) TEMP
                                         WHERE ROW_NUMBER = 1
                                     ),
```

The full query can be seen here
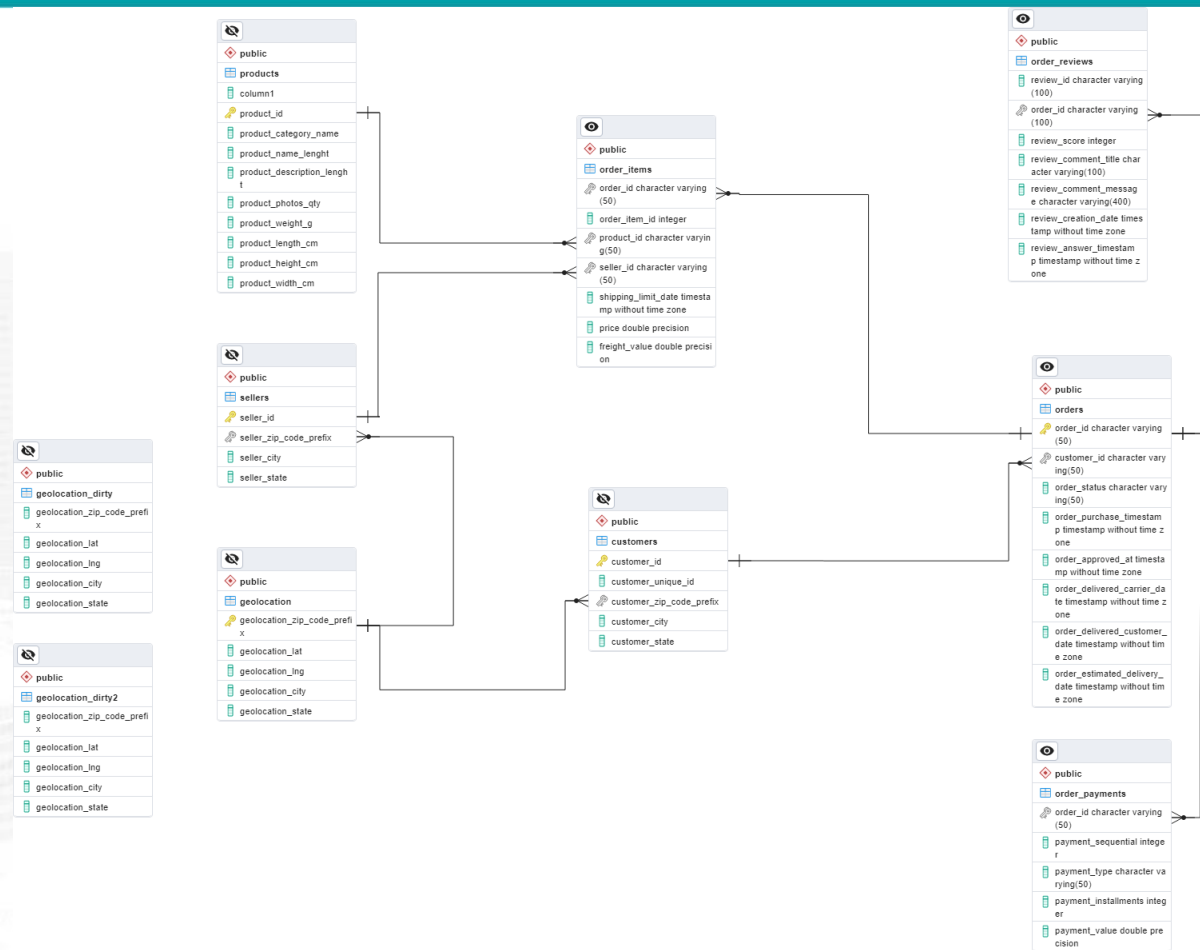
3. Entity Relationship :

Create an ERD by defining the Primary key and Foreign key for each table. Next, link the relationships between the keys by specifying the Constraint name. This results in the formation of an Entity Relationship Diagram (ERD) as illustrated in the picture.

```
-- customers -> orders

ALTER TABLE orders
ADD CONSTRAINT orders_fk
FOREIGN KEY (customer_id) REFERENCES customers(customer_id)
ON DELETE CASCADE ON UPDATE CASCADE;


-- geolocation -> customers

ALTER TABLE customers
ADD CONSTRAINT customers_fk
FOREIGN KEY (customer_zip_code_prefix) REFERENCES geolocation(geolocation_zip_code_prefix)
ON DELETE CASCADE ON UPDATE CASCADE;


-- geolocation -> sellers

ALTER TABLE sellers
ADD CONSTRAINT sellers_fk
FOREIGN KEY (seller_zip_code_prefix) REFERENCES geolocation(geolocation_zip_code_prefix)
ON DELETE CASCADE ON UPDATE CASCADE;
```

View or Download ERD here
The full query can be seen here

Annual Customer Activity Growth Analysis

1. Average Monthly Active User (MAU) per year

	year double precision 🔒	average_mau numeric 🔒
1	2016	108.67
2	2017	3694.83
3	2018	5338.20

2. Total new customers per year

	year double precision 🔒	new_customers bigint 🔒
1	2016	326
2	2017	43708
3	2018	52062

The full query can be seen [here](#)

AverageMonthly Active User and Total New Customers

— Average Monthly Active User (MAU) per year
— Total new customers per year
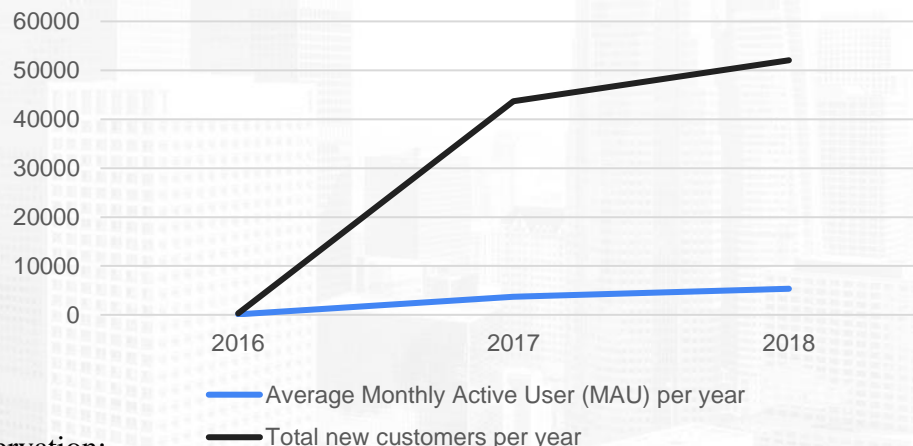
Observation:
● Monthly Active Users have consistently grown each year, accumulating thousands of customers.
● There was a notable surge in 2017, although this might be attributed to the data being accessible for only four months starting in September. Subsequently, it continued to rise in 2018.
● This indicates that e-commerce performance is correlated with strong engagement that captures customer interest, as evidenced by the upturn in the average active customer. Furthermore, effective branding promotion in e-commerce can attract new customers and encourage initial orders. Additionally, the possibility of offering promotions or vouchers to entice new customers is worth considering.
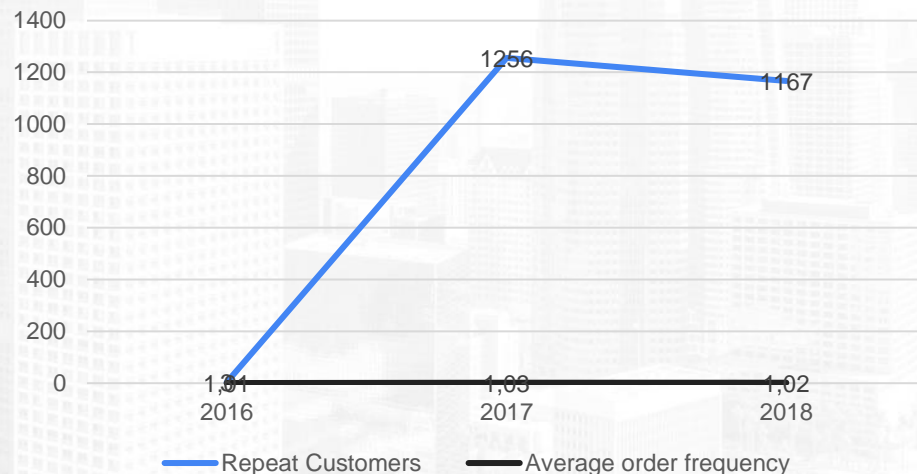
3. The number of customers who make repeat orders per year

	year double precision 🔒	repeat_customers bigint 🔒
1	2016	3
2	2017	1256
3	2018	1167

4. Average order frequency for each year

	year double precision 🔒	avg_num_orders numeric 🔒
1	2016	1.01
2	2017	1.03
3	2018	1.02

Repeat Customer and Avg Order Freq

Observation:
● There was a substantial rise in repeat orders in 2017. Nevertheless, there was a decline in 2018, amounting to 100 customers.
● By examining the average order quantity, it becomes apparent that the majority of individuals placed just one order in the past three years.

The full query can be seen here
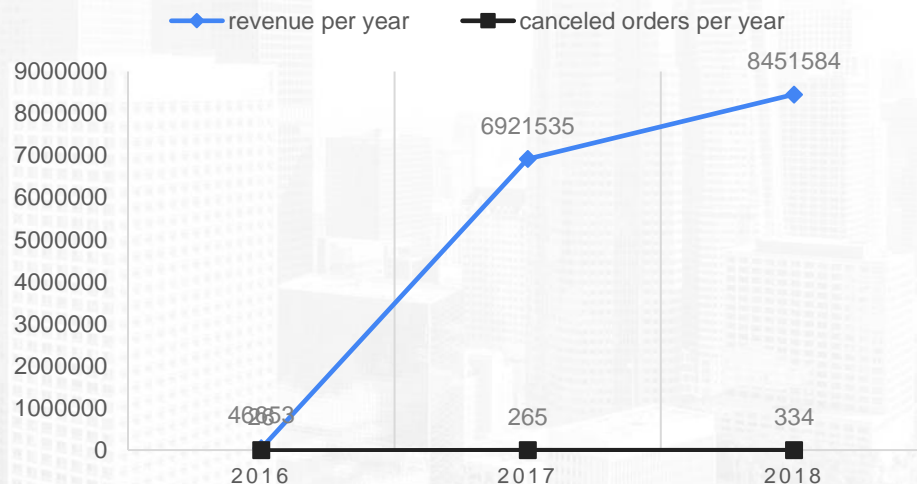
Rakamin Academy

1. Table of revenue per year

	year double precision 🔒	revenue double precision 🔒
1	2016	46653.74000000004
2	2017	6921535.239999719
3	2018	8451584.769999959

2. Table of the number of canceled orders per year

	year double precision 🔒	total_cancel bigint 🔒
1	2016	26
2	2017	265
3	2018	334

The full query can be seen here

REVENUE AND CANCELED ORDERS PER YEAR

Observation:
● Revenue has been consistently increasing every year, despite the previous data indicating a decline in repeat orders from 2017 to 2018. This trend may be attributed to a significant rise in the number of new customers each year.
● The number of canceled orders increased from 2016 to 2018, although the scale of the increase remained relatively small, still in the hundreds as opposed to the thousands. It would be advisable to gather additional data on the reasons for order cancellations to prevent a further rise in cancellations in the upcoming year.

3. Table of the top categories that generate the largest revenue per year

	year double precision	product_category_name character varying	revenue double precision
1	2016	furniture_decor	6899.35
2	2017	bed_bath_table	580949.200000002
3	2018	health_beauty	866810.3399999985

4. Table of categories that experienced the most canceled orders per year

	year double precision	product_category_name character varying	total_cancel bigint
1	2016	toys	3
2	2017	sports_leisure	25
3	2018	health_beauty	27

Observation:
- It is apparent that the various product categories, including the ones with the highest revenue and those with the most cancellations, differ every year.
- Nevertheless, an intriguing discovery was made in 2018 when it was observed that the Health Beauty category held the same position for both the highest revenue and the most cancellations. This might be attributed to the fact that a significant number of orders were placed in the health and beauty categories.

The full query can be seen here

1. Total usage of each type of payment all time

| payment_type 🔒 character varying	num_of_usage 🔒 bigint
1	credit_card	76795
2	boleto	19784
3	voucher	5775
4	debit_card	1529
5	not_defined	3

Payment Type Used for ALL Time
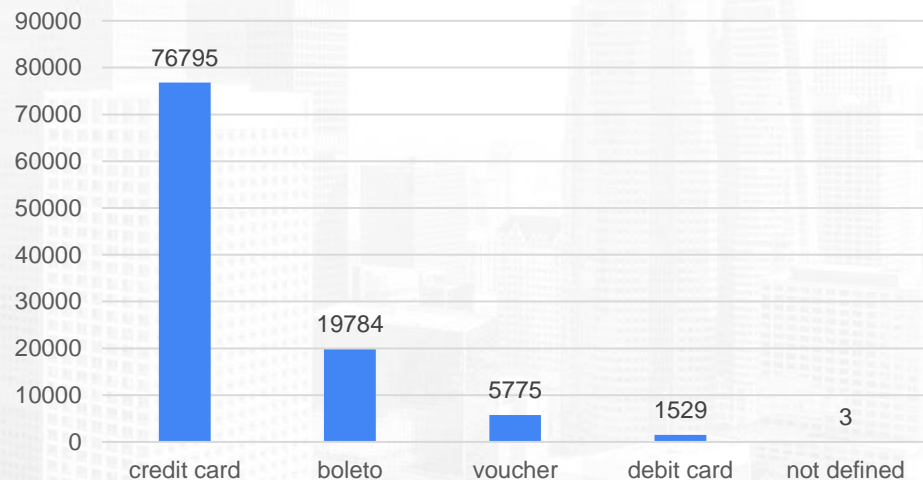
Observation:
● There are 4 common payment methods, which are Credit Card, Boleto, Voucher, Debit Card.
● The use of credit cards as a means of payment is the most compared to other types of payments. By looking at this, companies can collaborate with credit card provider banks to offer more appealing promotions. Further analysis can be conducted on customer behavior in using credit cards, such as the chosen credit type, which product categories are usually purchased with credit cards, etc.
● Because many customers use credit cards, boleto, and vouchers, bonuses or promotions can be provided that target the most favorite payment types.

The full query can be seen here

2. Details of the amount used for each type of payment for each year

	payment_type character varying	year_2016 numeric	year_2017 numeric	year_2018 numeric
1	credit_card	258	34568	41969
2	boleto	63	9508	10213
3	voucher	23	3027	2725
4	debit_card	2	422	1105
5	not_defined	0	0	3

Payment Type for 2016,2017 and 2018
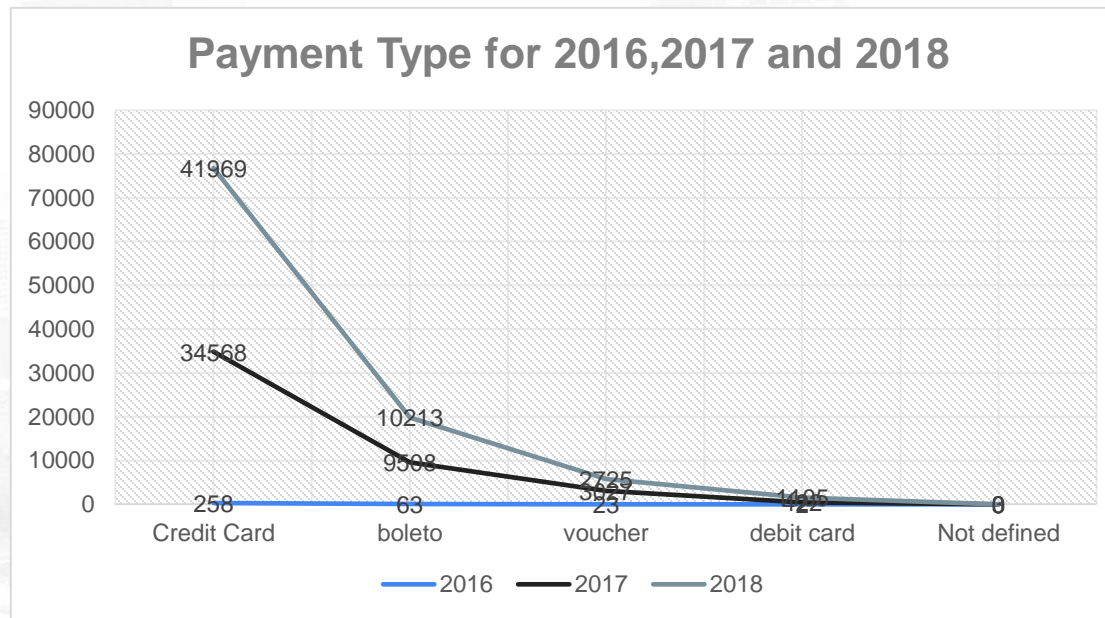
Observation:
- Additionally, each payment type exhibits a notable annual increase.
- Nevertheless, voucher-based payments experienced a decline in 2018, possibly attributed to reduced voucher promotions.
- Further examination can be conducted by consulting with other departments, including Marketing, Analysis, or Business Development.

The full query can be seen here

1. Customer activity experienced growth in every metric from 2016 to 2018, including an increase in the number of new customers and Monthly Active Users (MAU). However, repeat customer orders remained stagnant during this period. Moreover, the average number of orders made by customers was only once.

2. The analysis of Product Category Quality revealed a consistent growth in the company's total revenue each year. Interestingly, the most canceled product categories and best selling orders changed annually. Notably, the health beauty category emerged as both the most sold and the most canceled product category.

3. Additionally, each type of payment showed a significant increase year over year. Credit cards were the most commonly used payment method from 2016 to 2018. However, payments made using vouchers experienced a decline in 2018.