

# Analyzing eCommerce Business Performance with SQL

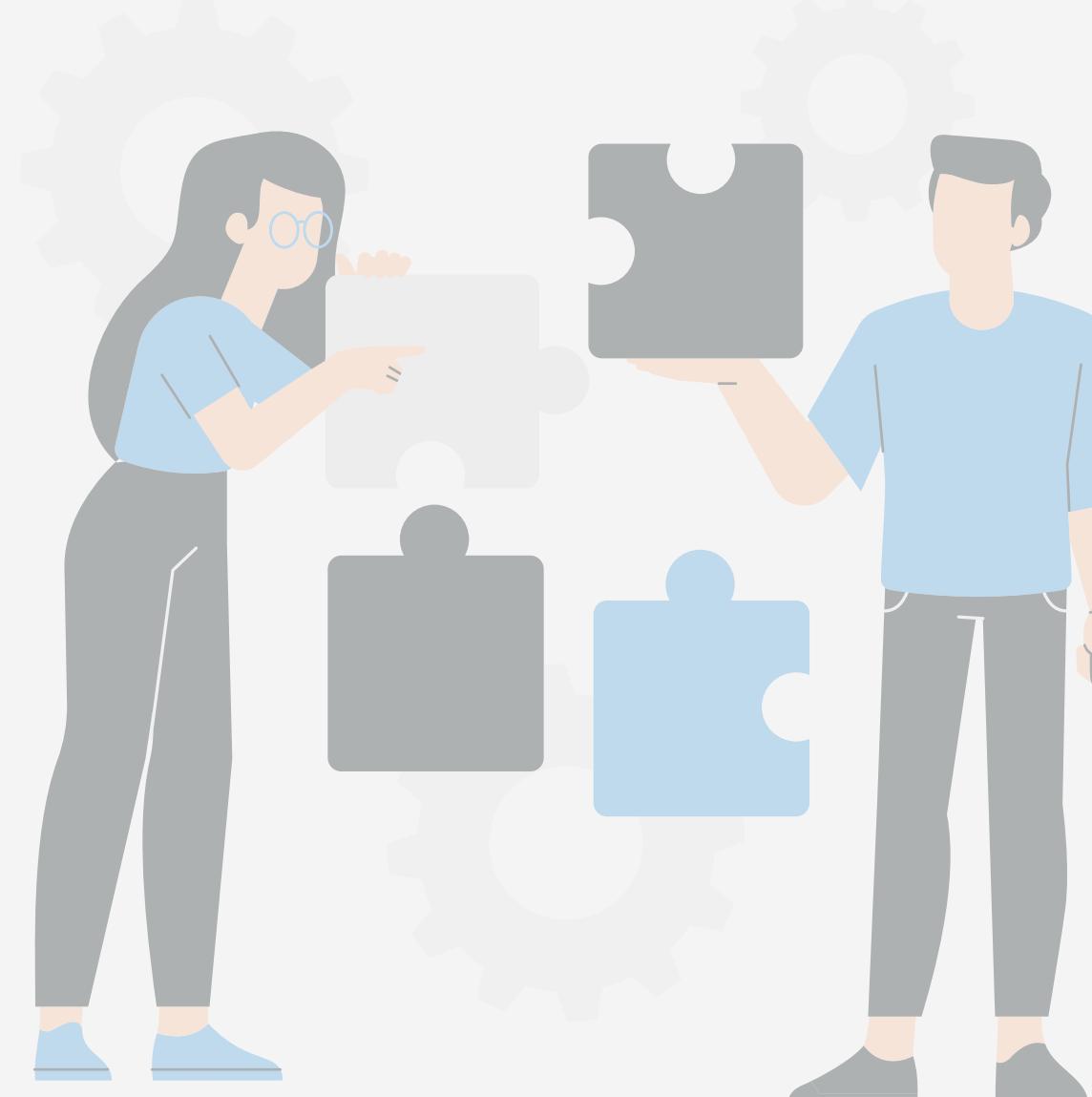
**Supported by:**  
**Rakamin Academy**  
Career Acceleration School  
[www.rakamin.com](http://www.rakamin.com)



**Created by:**  
**Lutfia Husna Khoirunnisa**  
Your Email:  
[lutfiahusnakhoirunnisa@gmail.com](mailto:lutfiahusnakhoirunnisa@gmail.com)  
Your linkedIn Profile  
[linkedin.com/in/lutfiahusnakhoirunnisa](https://linkedin.com/in/lutfiahusnakhoirunnisa)

**“Lutfi is a junior data analyst experienced on data analysis, business analysis, and data science with a background in mathematics. Experienced in handling and interpreting diverse data sets, extracting valuable insights, and making datadriven recommendations ”**

# Step by Step



1

Project Introduction

2

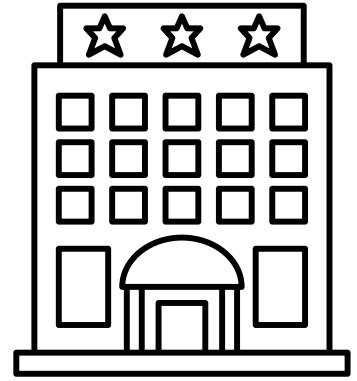
Data Preparation

3

Data Analysis

4

Conclusion



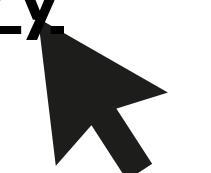
# Project Introduction

In a company, measuring business performance is crucial for tracking, monitoring, and assessing the success or failure of various business processes. Therefore, this project will analyze the business performance of an eCommerce company, taking into account several business metrics, namely customer growth, product quality, and payment methods.



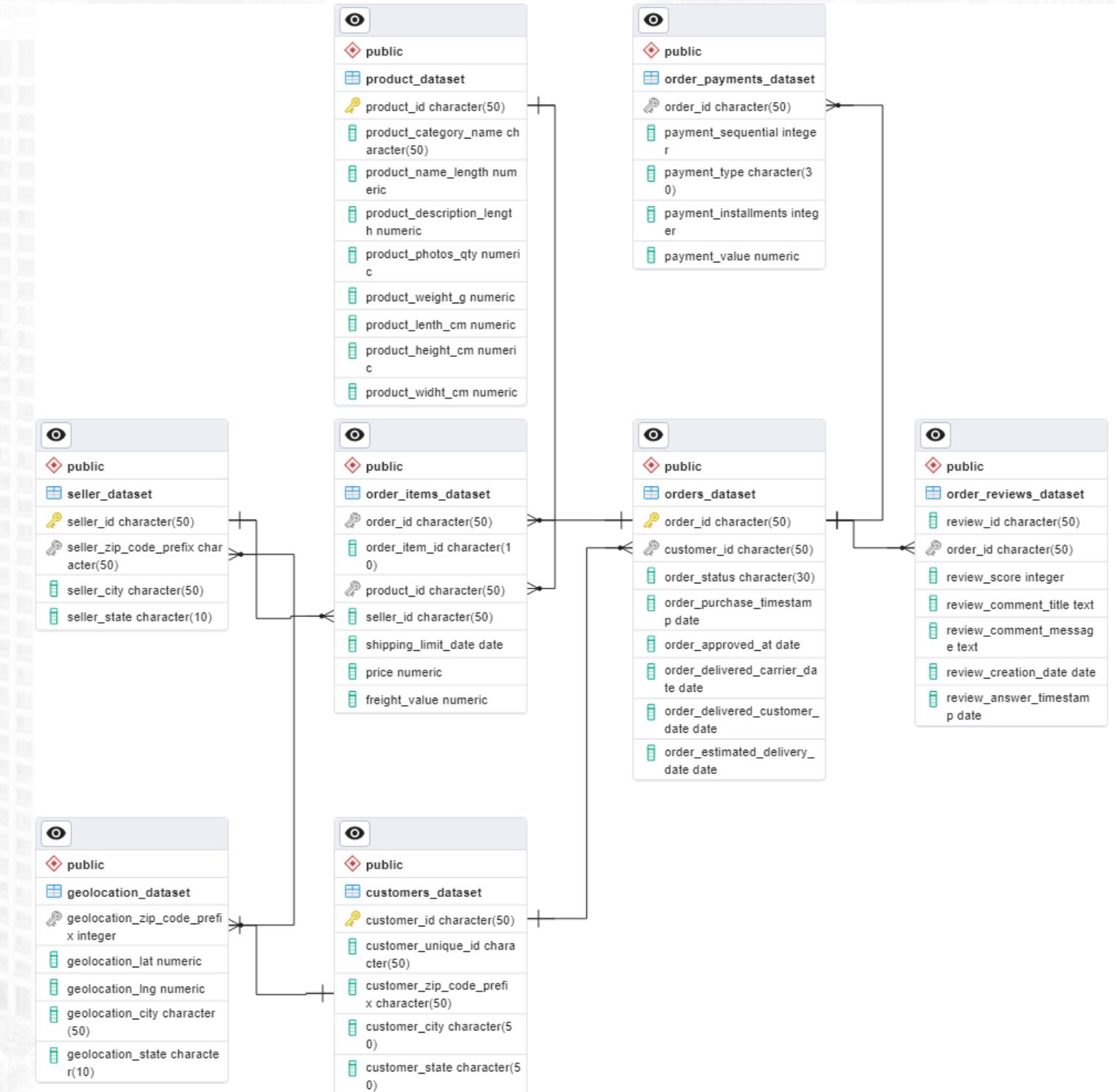
Tools : Postgre-SQL,  
Excel & Canva-Data Visualization

[Click here to access the query  
and dataset](#)



# Data Preparation

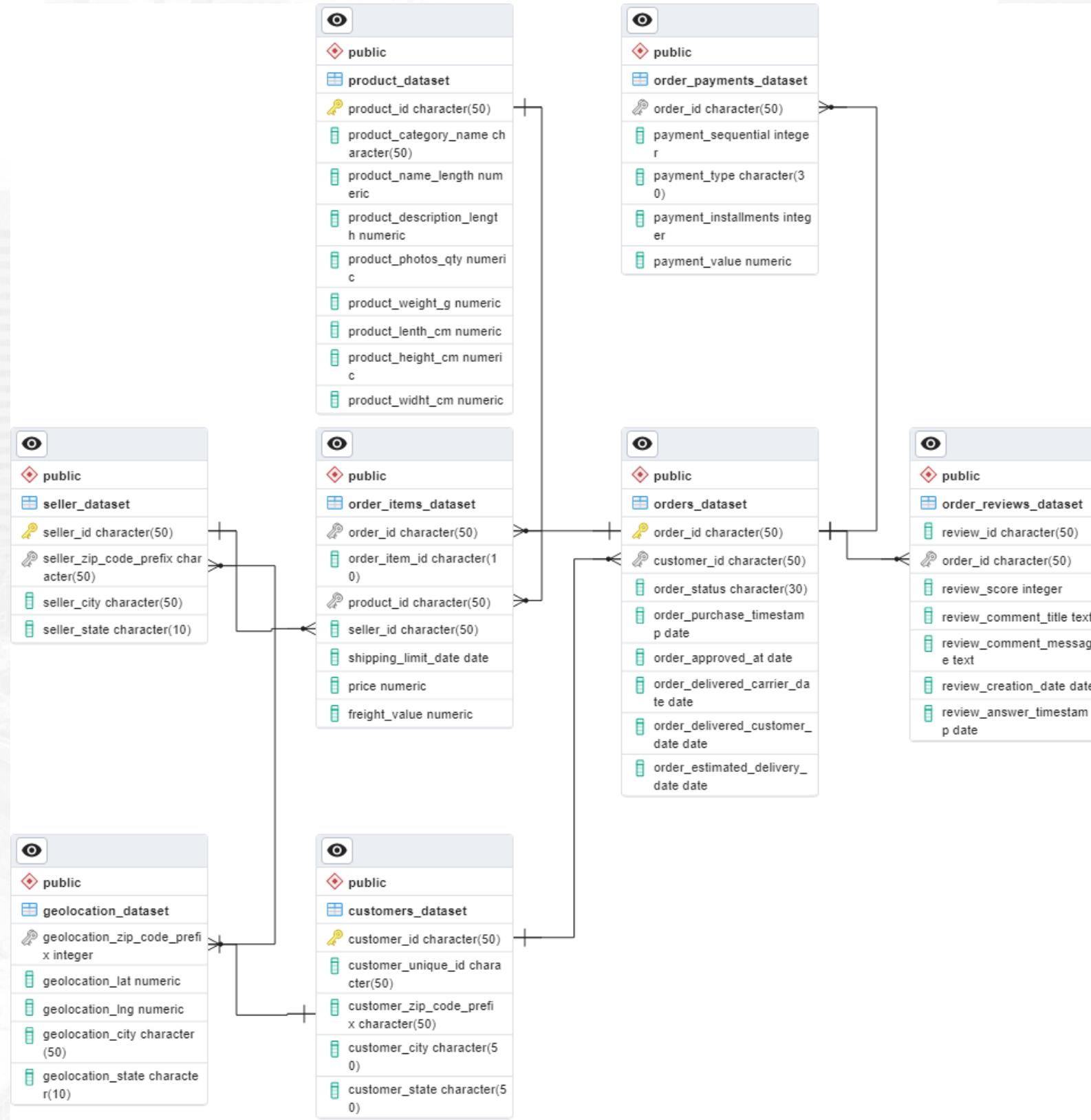
- Data preparation begins by creating tables and importing the data into them.
- An Entity-Relationship Diagram (ERD) is created to illustrate the relationships between the datasets.
- All the data preparation work is performed using PostgreSQL and pgAdmin.



# Data Preparation

## HUBUNGAN ANTAR DATASET DIJELASKAN DALAM ERD

BERIKUT :



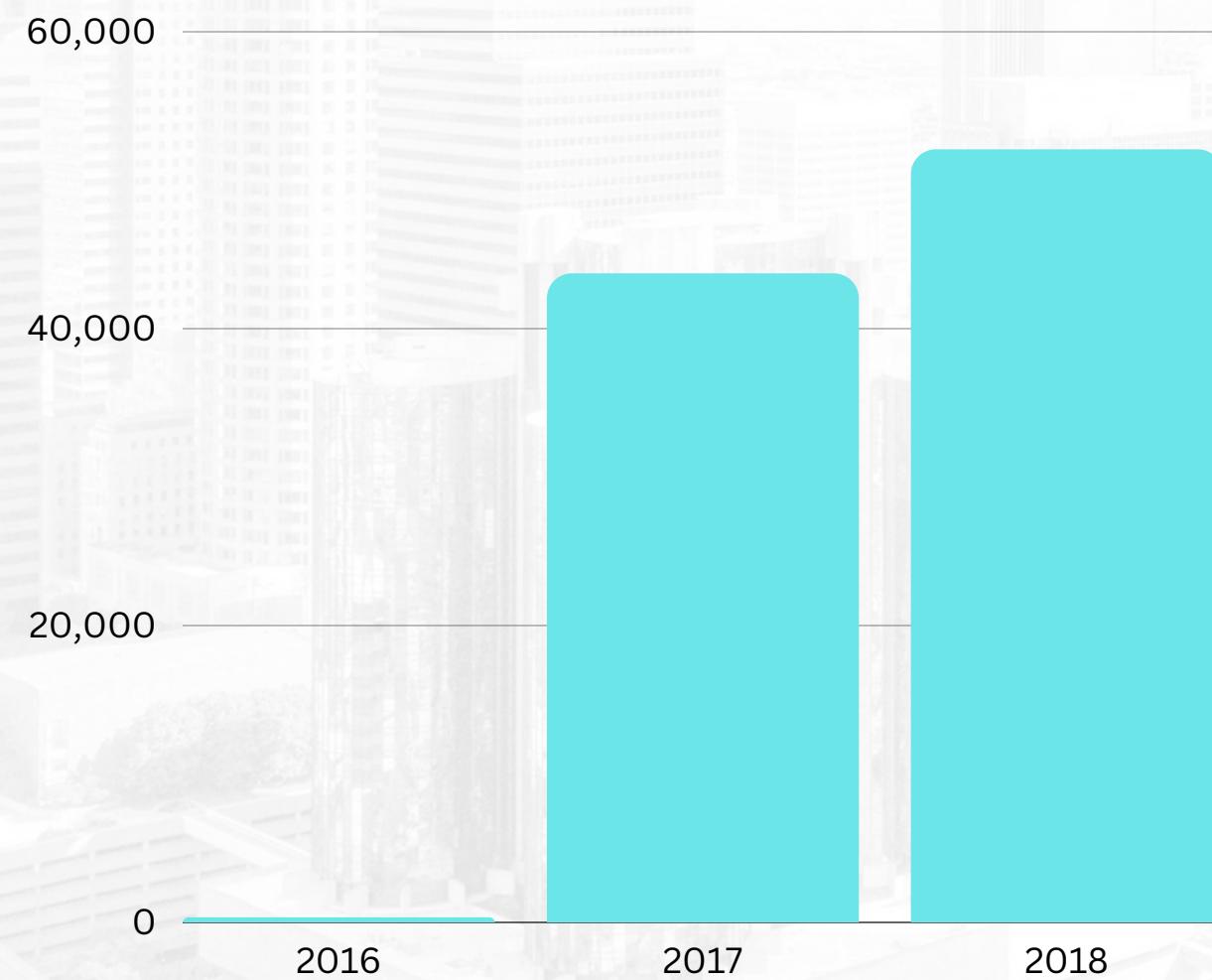
The datasets used are as follows:

- **Sellers:** Contains information about sellers.
- **Customers:** Contains information about customers.
- **Geolocation:** Contains geolocation information of customers and sellers, such as latitude and longitude coordinates, zip\_code\_prefix key.
- **Product:** Contains information about the products being sold, such as product name, description, weight, length, etc.
- **Order\_items:** Contains information about the purchased items, including order\_id, seller ID, product ID, and other details such as price, etc.
- **Orders:** Contains detailed information about the orders, including order\_id (which is used to link with the order\_items dataset), customer\_id, status, purchase date, etc.
- **Order\_payments:** Contains information about the payments made, including order\_id, total payment, payment method, etc.
- **Order\_reviews:** Contains information about the reviews submitted by customers.

# Data Analysis

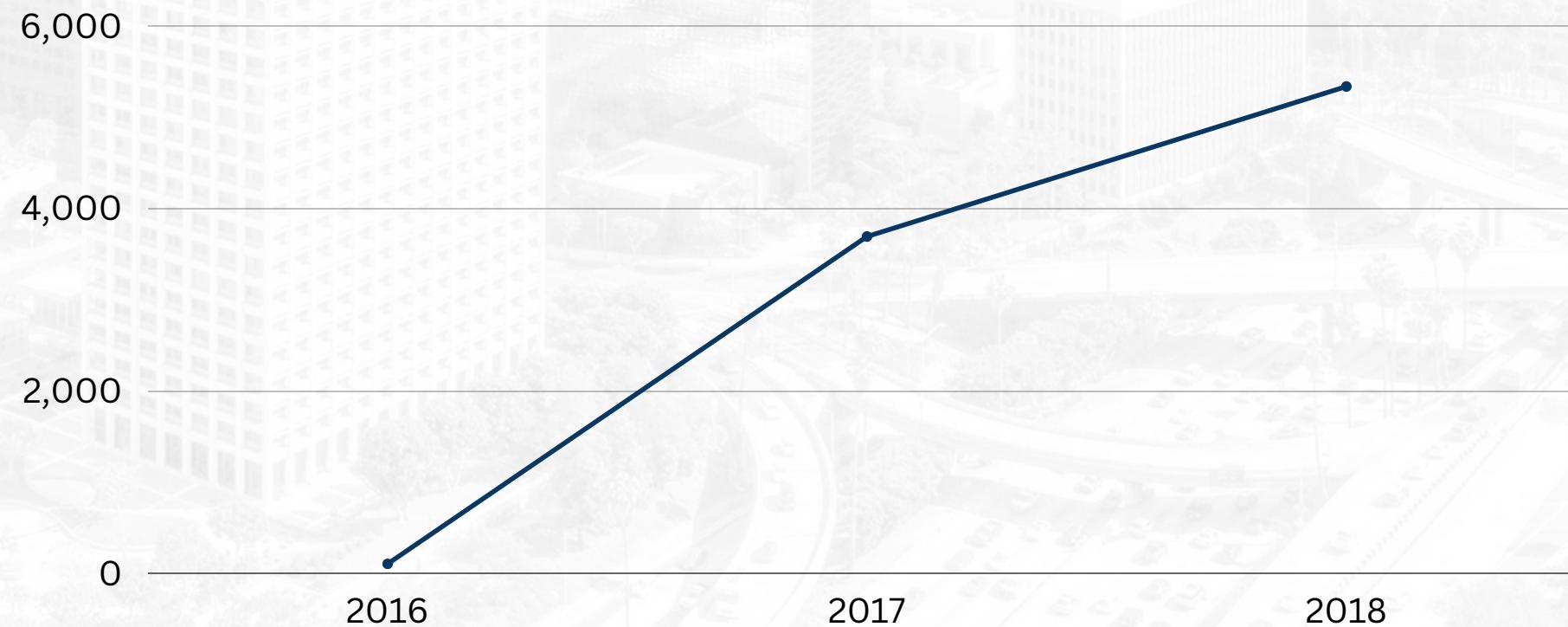
## (Annual Customer Activity Growth Analysis)

NEW CUSTOMER BY YEAR



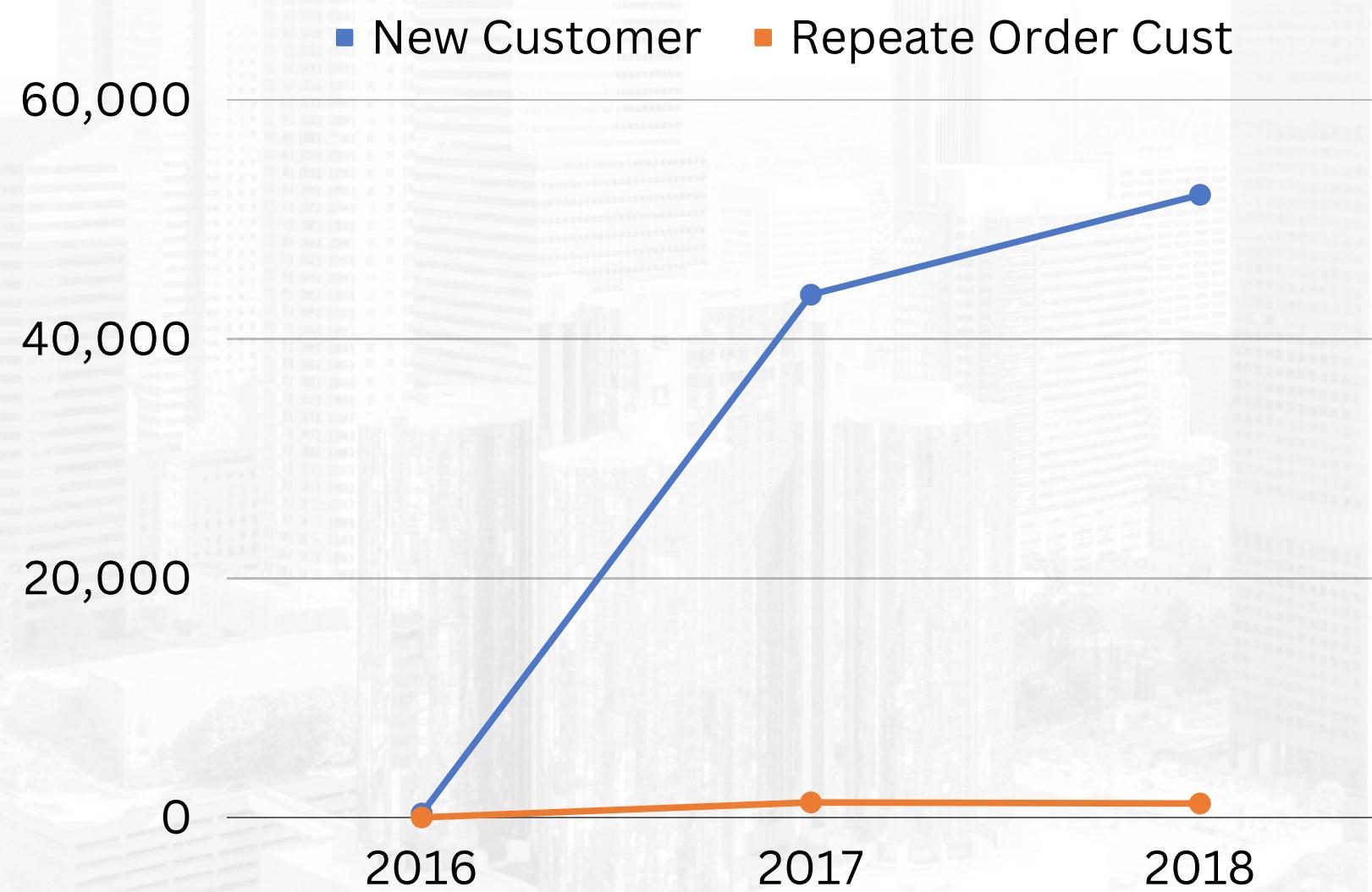
The number of new customers increases each year, and this is also consistent with the continuously increasing average monthly active users from 2016 to 2018.

AVERAGE MONTHLY ACTIVE USER



# Data Analysis

## (Annual Customer Activity Growth Analysis)



Although the numbers of new customers and average monthly active users appear promising each year, it is important to consider that the total number of customers who make repeat orders annually remains significantly low compared to the growth of new customers.

Table of New Customer and Repeat Customer Numbers per Year

active_year	new_cust	tot_repeat_cust
2016	326	3
2017	43708	1256
2018	52062	1167

**"The company needs to conduct further evaluation to analyze the cause of the low customer repeat order issue."**

# Data Analysis

## (Annual Customer Activity Growth Analysis)

From the ratio of new customers to customers who make repeat orders, it is known that the **company has a serious issue in terms of customer loyalty**. Based on the table provided, it is observed that on average, customers only make one transaction per year.

The company needs to develop strategies to maintain or attract customers to make repeat transactions. Several actions need to be taken, including:

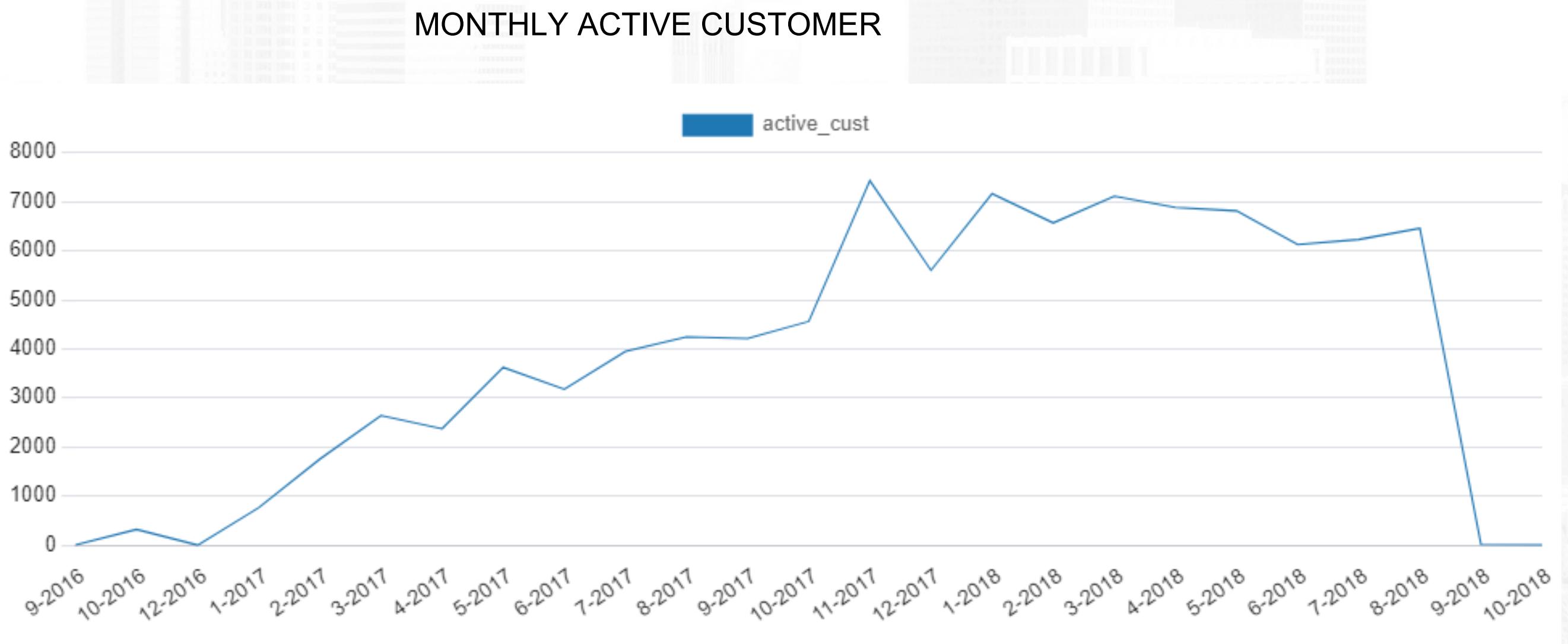
- Analyze customer experience to identify any differences between new customers and customers who make repeat orders.
- While acquiring new customers is important, it is equally crucial to focus on customer retention. Campaigns or promotions can be implemented to re-engage customers and regain their interest.

active_year double precision	avg_year_freq numeric
2016	1.01
2017	1.03
2018	1.02

The average frequency of customer transactions per year

# Data Analysis

## (Annual Customer Activity Growth Analysis)



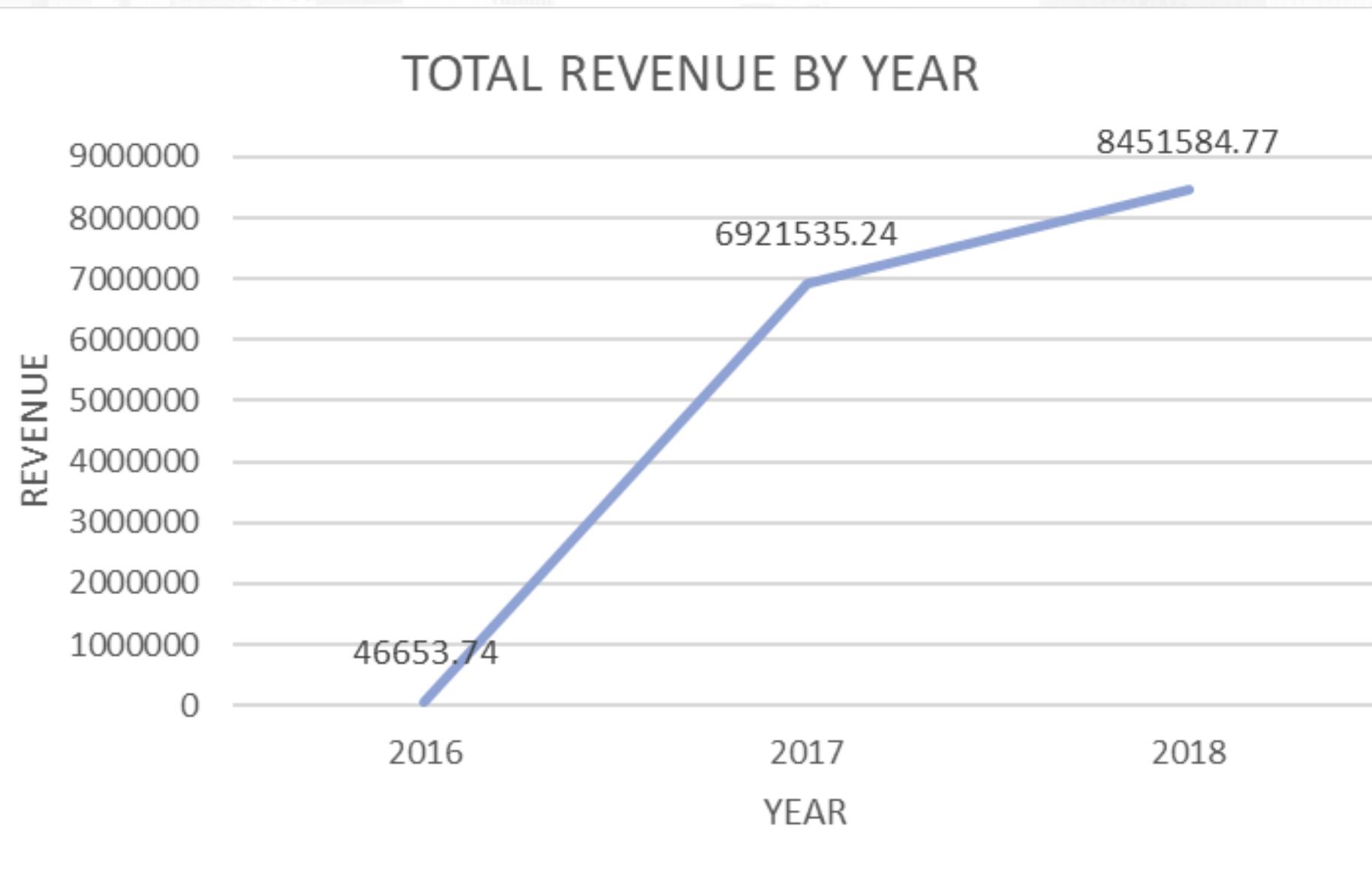
**"The number of active customers dropped significantly at the end of 2018. It appears that the company is facing a serious issue in customer retention."**

### Insights:

- The total number of active customers in 2017 tends to be the most stable compared to 2016 and 2018.
- In 2016 and 2017, there is a significant decrease of the total number of active customers at the end of the year, followed by an increase at the beginning of the year.
- In September 2018, there is an extreme decrease of active customers, and the decline continues in the following months.

# Data Analysis

## (Annual Revenue Analysis)

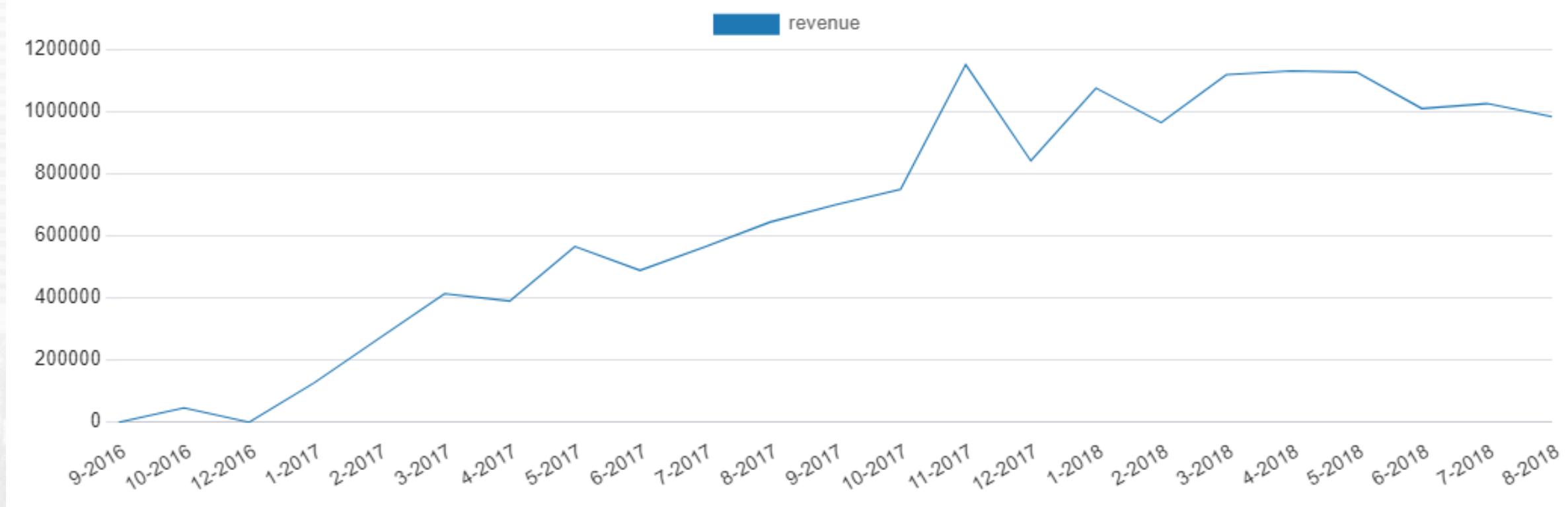


The revenue in 2016 was still very low, but this was due to sales that only occurred in the last 4 months of the year.

There is an increase in total revenue each year, and the revenue in 2018 is the highest compared to the previous 2 years. However, based on the previous analysis that revealed the Monthly Active Users extremely dropped the end of 2018, we need to conduct further analysis on the total revenue to determine if there is any impact from the drastic decline in Monthly Active Users.

# Data Analysis

## (Annual Revenue Analysis)



In general, the revenue appears to improve year by year, even though the revenue decrease at the end of each year, but it increase again at the beginning of the following year.

In accordance with the fact that the extreme decrease of Monthly Active Users in September, the revenue also stopped in after August 2018.

The table beside shows the customer order activity based on their order status after August 2018. There are only orders with the status "canceled" and "shipped," indicating that no revenue was generated after August.

order_status	count
canceled	19
shipped	1

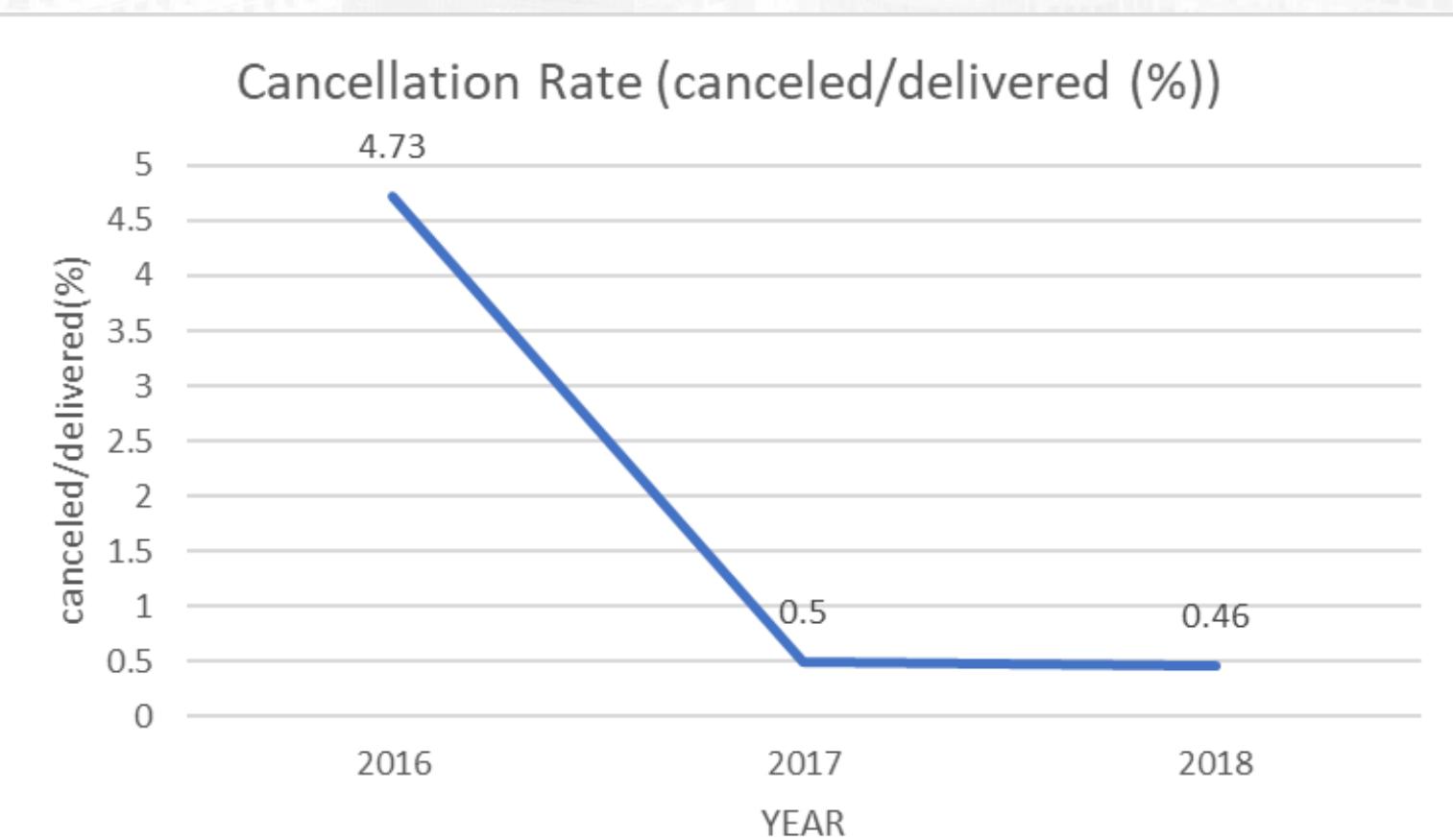
# Data Analysis

## (Annual Order Quality Analysis)



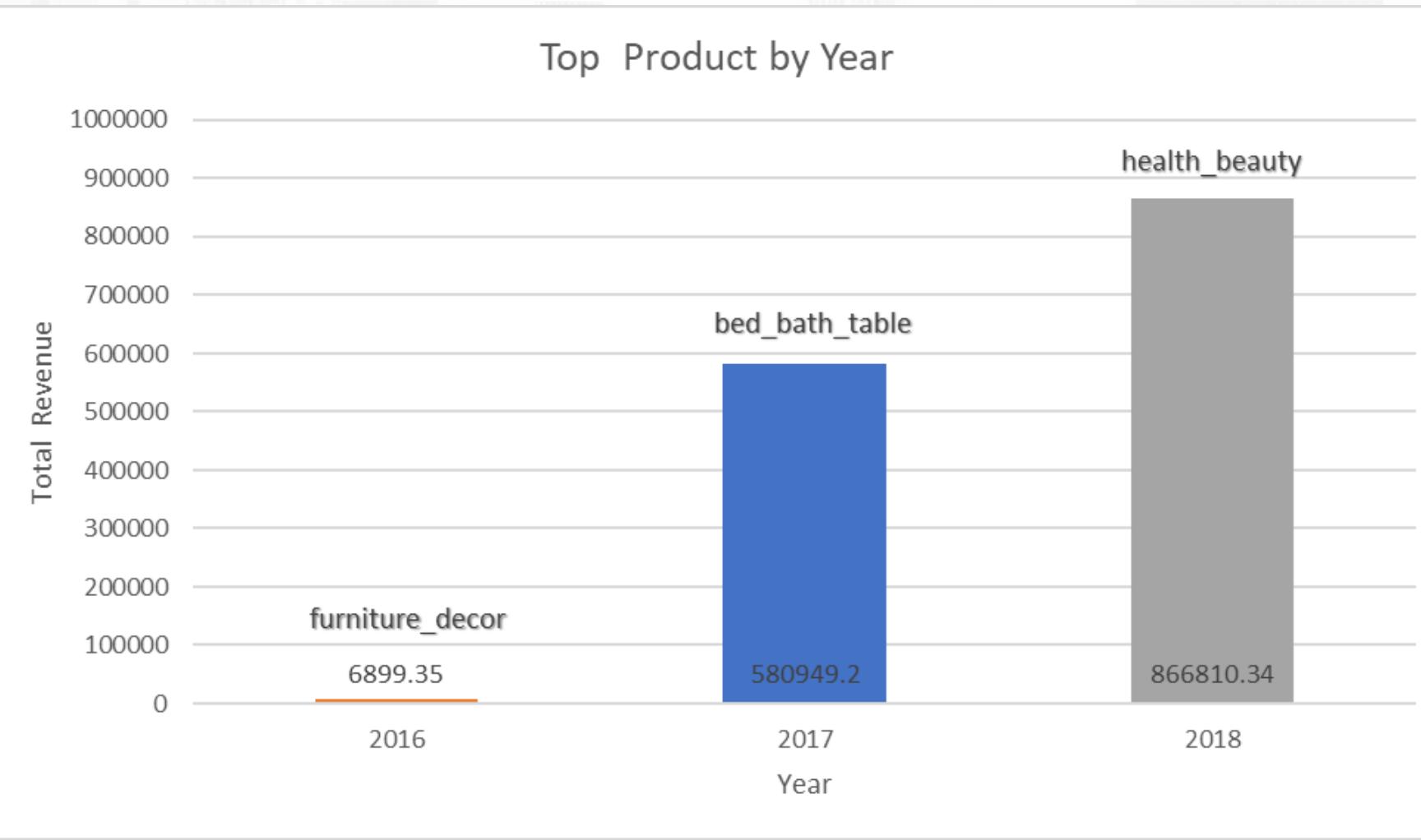
However, when considering the cancellation rate, it turns out that 2017 had the highest cancellation rate.

The number of canceled orders keeps increasing every year, reaching its peak in 2018.

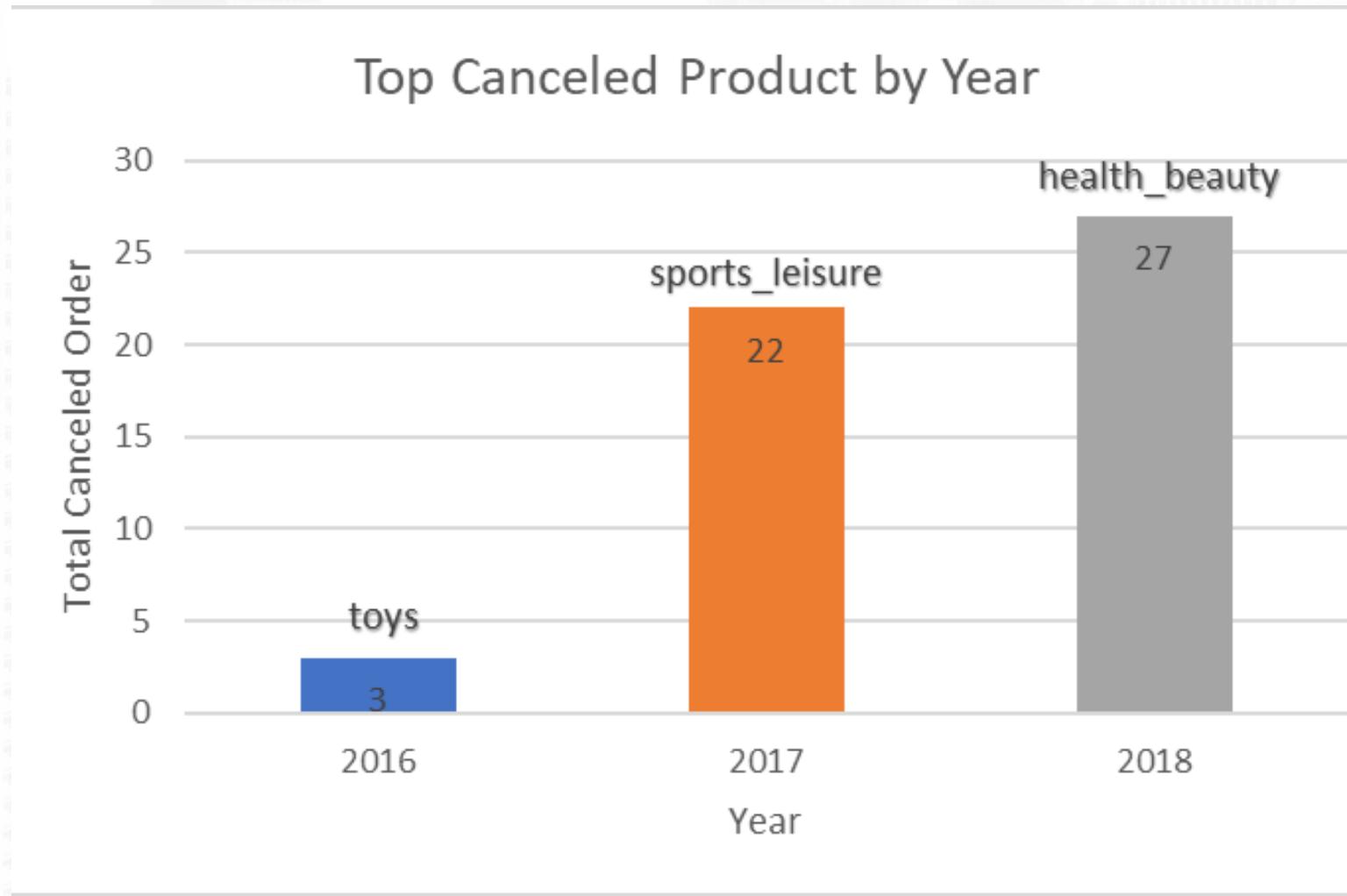


# Data Analysis

## (Annual Product Category Quality Analysis)



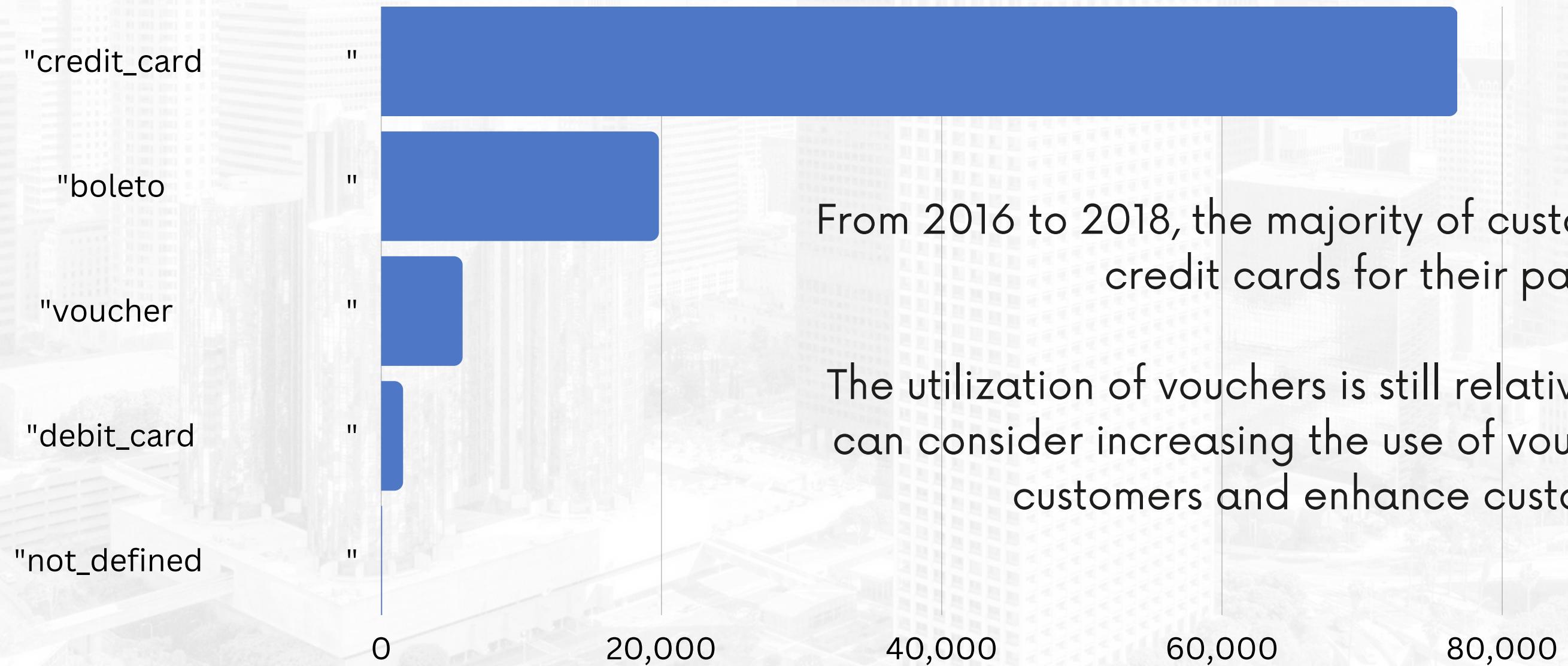
The top products differ each year. By considering the total revenue generated for each product category, we can identify the flagship products for each year in the graph above.



The products with the highest number of canceled orders also differ each year. When examining the total canceled orders for each product category, we can observe the products with the highest canceled orders for each year in the graph above.

# Data Analysis

## (Analysis of Annual Payment Type Usage)

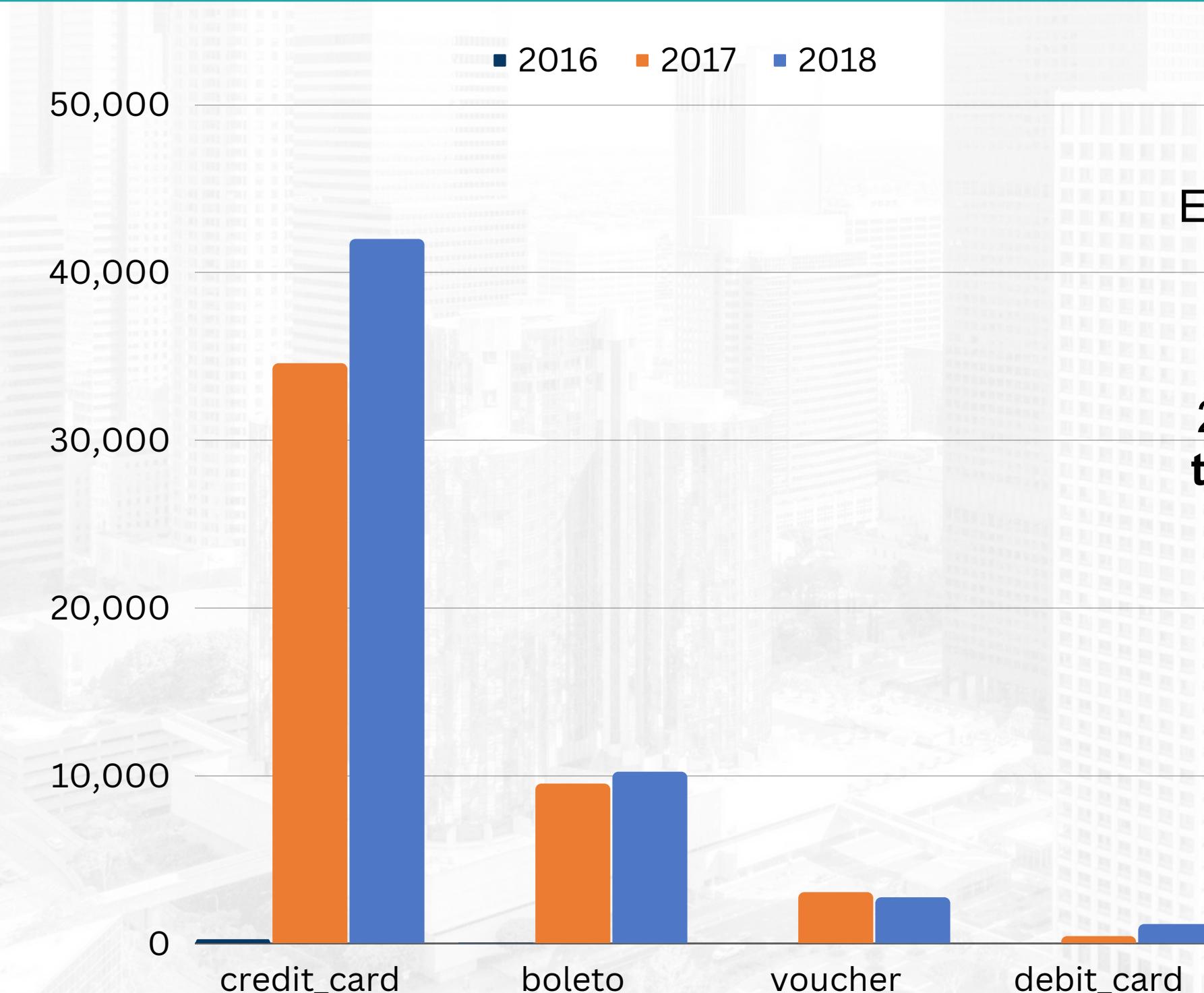


From 2016 to 2018, the majority of customers preferred to use credit cards for their payments.

The utilization of vouchers is still relatively low. The company can consider increasing the use of vouchers to attract more customers and enhance customer loyalty.

# Data Analysis

## (Analysis of Annual Payment Type Usage)

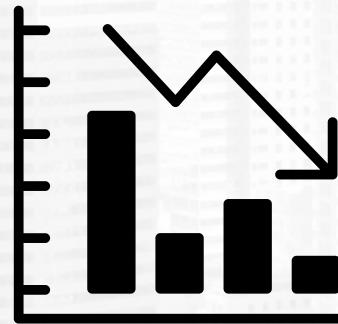


Each year, the most common payment method is credit card, and its usage frequency increases annually.

The usage of vouchers significantly increased from 2016 to 2017 and dropped in 2018. This indicates that **there might have been some issues with vouchers in 2018**, such as limited availability or ineffective marketing communication to customers, especially considering the majority of customers were new customers.

# Conclusion

Further analysis is needed to determine the reasons behind the drastic decline in sales at the end of 2018. Several factors should be focused:



The revenue in August 2018 requires further investigation to identify any events or factors that may have caused the dropped sales.



Customer experience analysis is necessary to understand why many customers are not making repeat orders. It is also important to identify the reasons behind the high number of order cancellations.



The company needs to prioritize improving customer loyalty, as the number of customers making repeat orders is significantly low and disproportionate compared to the number of new customers.



The decrease in voucher usage in 2018 needs to be thoroughly analyzed to determine the underlying causes. This includes examining whether there was a shortage of available vouchers or there was marketing problem to inform customers about the existence of vouchers, especially considering the majority of customers are new customers.

## **Mini Project**

Completion Certification

has been presented to

# **LUTFIA HUSNA KHOIRUNNISA**

For successfully completing data scientist  
Mini Project Analyzing eCommerce Business Performance with  
SQL

CEO Rakamin Academy



**Andika Deni Prasetya**



Data Scientist,  
Advance AI



**Gerry Chandra**

# **Thank you**

**Check out my profile and  
other portfolios:**

 [Linkedin](#)

 [GitHub](#)