

ANALYZING ECOMMERCE BUSINESS PERFORMANCE WITH SQL



Presented By
Devina

Hello, I am Devina

A Data and Python Enthusiast



A self-motivated person with integrity who is always eager to develop my skills to grow both personally and professionally. I can work efficiently either individually or with a team. A multi-tasker, detail-oriented person, and responsible.”

Education: Bachelor Degree of Environmental Engineering at University of Sumatera Utara

Experiences: Sustainable Supply Chain – Traceability Data & System (Feb 2020 – Aug 2021)

Tools: Python, PostgreSQL, Microsoft Office

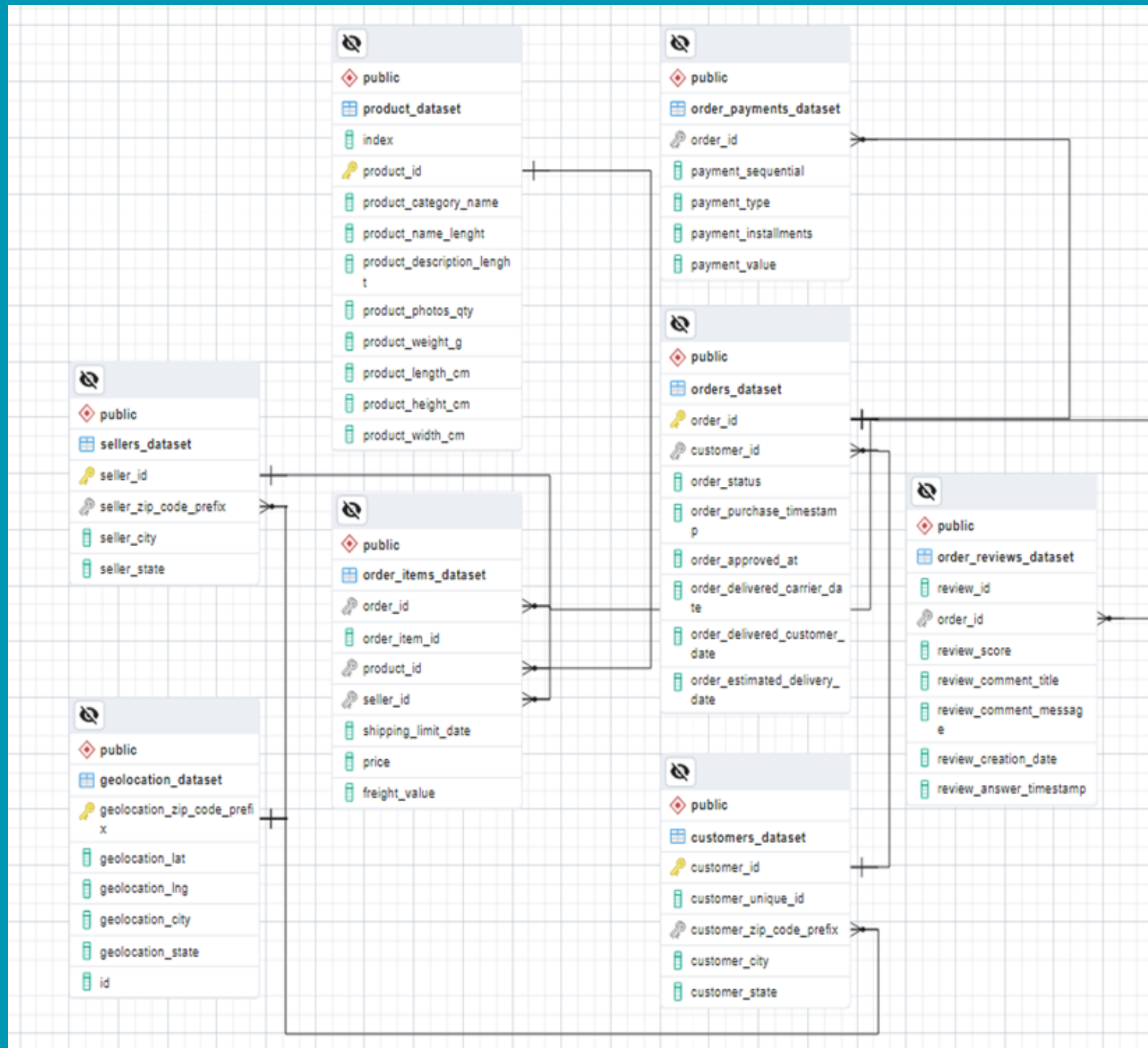
Skills: Analytical Skills, Time Management, Data Entry, Exploratory Data Analysis, Machine Learning Modelling and Evaluation

Introduction

It is very important to track, monitor, and judge the success or failure of any business process in a company. Therefore, this paper will analyze the business performance of an ECommerce company by considering a few metrics, such as customer growth, product quality, and payment type.



Entity Relationship Diagram



Entity Relationship Diagram (ERD) Summary:

- **order_id** on **orders_dataset** as the Foreign Key for **order_payments_dataset**, **order_reviews_dataset**, and **order_items_dataset**.
- **customer_id** on **customers_dataset** as the Foreign Key for **orders_dataset**.
- **product_id** on **product_dataset** as the Foreign Key for **order_items_dataset**.
- **seller_id** on **sellers_dataset** as the Foreign Key for **order_items_dataset**.
- **geolocation_zip_code_prefix** on **geolocation_dataset** as the Foreign Key for **sellers_dataset** and **customers_dataset**.

Check out the ERD [here](#).

Check out the queries [here](#).

Data Preparation

Steps during the Data Preparation:

1. Review the CSV files and create the tables by the available datasets (8). Ensure that every column on each label already has the appropriate data type set.
2. Import the dataset to the created tables (import CSV).
3. Analyze the relationship between each table and set the Primary Key for each table (if available) by giving the tick on the "**Primary Key?**" option.
4. Add the Foreign Key if available by using the *alter table* command and manually by using the option available in the table's **Properties**, then open the **Constraints**, and add Foreign Key data (This step was done on **sellers_database** and **customers_database** while adding the Foreign Key for the **zip_code_prefix** column).

Check out the ERD [here](#).
Check out the queries [here](#).

Data Preparation

Summary:

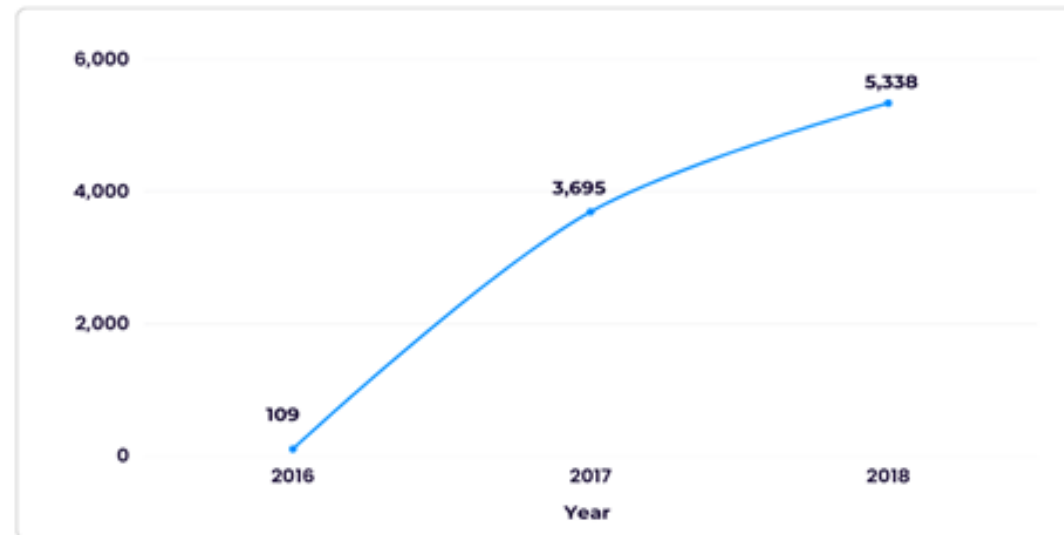
1. Check the received dataset (CSV) and create a database in accordance with the available dataset.
2. Connect the suitable tables by setting the **Primary Key** and **Foreign Key** between tables.
 - a. Delete the duplicates on the 'geolocation_zip_code_prefix' column which will be used as the primary key on the 'geolocation_dataset' table (the duplicates have similar core information with only slight difference on the 'geolocation_lat' and 'geolocation_lng' columns).
 - b. It's better if there is an update done on the 'geolocation_zip_code_prefix' column by adding the zip codes registered on the 'customer_zip_code_prefix' column (**customers_dataset** table) and 'seller_zip_code_prefix' (**sellers_database** table) but still unregistered on the 'geolocation_zip_code_prefix' on the **geolocation_dataset** table.
3. Create the Entity Relationship Diagram (ERD)

Check out the ERD [here](#).

Check out the queries [here](#).

Annual Customer Activity Growth Analysis

Average Monthly Active User



Monthly Active User

The average monthly active user increased significantly since year 2016 to the year 2018 where:



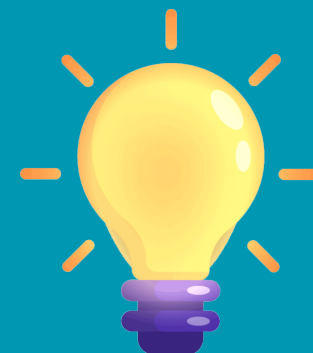
- Increase in the average monthly active user of 3568 users from the year 2016 to the year 2017.
- Increase in the average monthly active user of 1643 users from the year 2017 to the year 2018.

Yearly New Customer Growth



New Customer

New customer growth has significant increase from the year 2016 to year 2018 where:



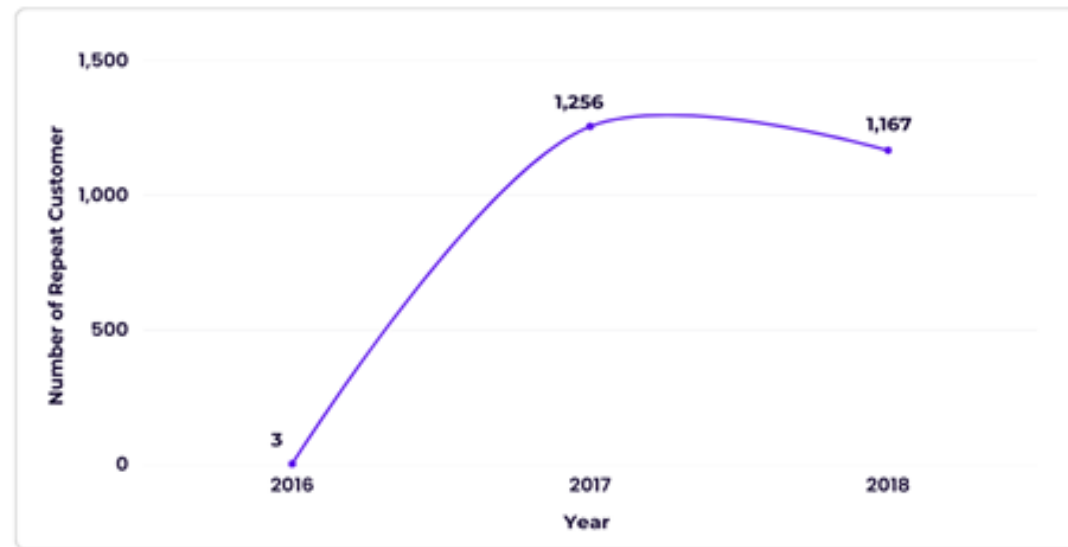
- Increase in new users of 43382 users from the year 2016 to the year 2017.
- Increase in new users of 8354 users from the year 2017 to the year 2018.

Check out the ERD [here](#).

Check out the queries [here](#).

Annual Customer Activity Growth Analysis

Yearly Repeat Customer Growth



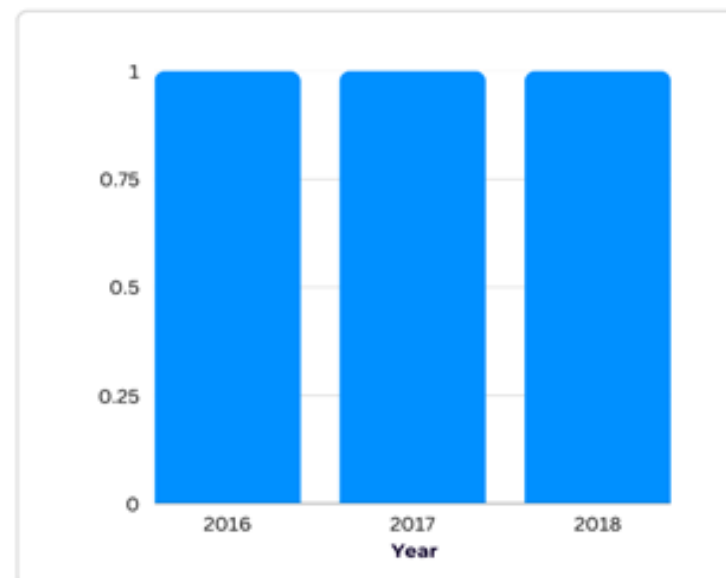
Repeat Customer



The repeating customers' growth has increased and decreased from the year 2016 to the year 2018, where:

- Increase of the repeating customers of 1253 users from the year 2016 to the year 2017.
- Decrease of the repeating customers of 89 users from the year 2017 to the year 2018.

Yearly Average Order Frequencies



Average Order Frequencies



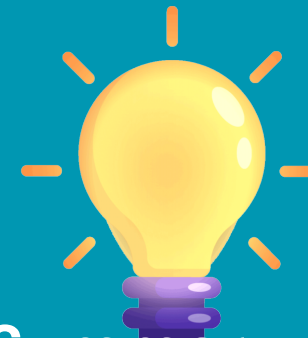
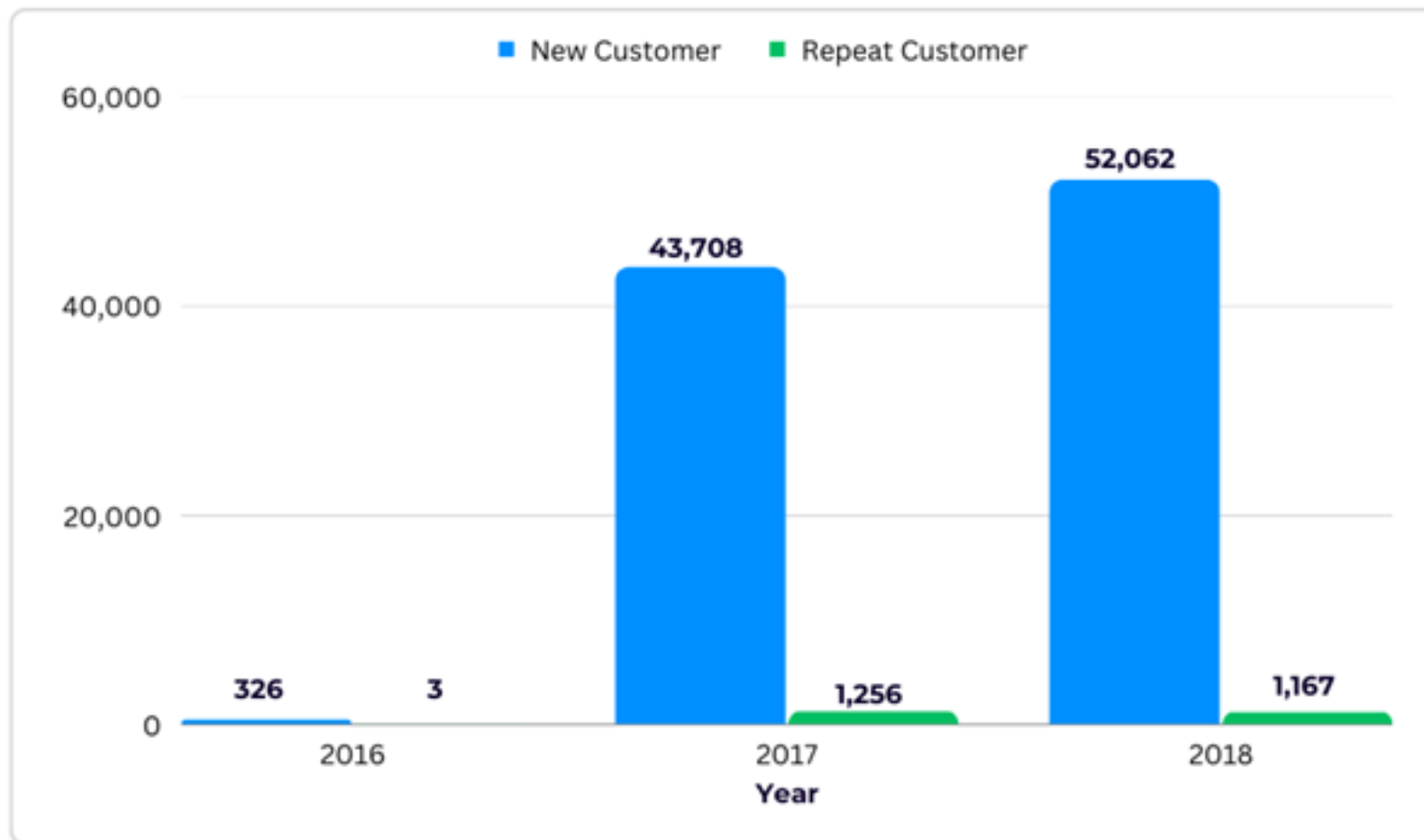
There was no significant increase or decrease in the average order frequencies every year from the year 2016 to the year 2018.

Check out the ERD [here](#).

Check out the queries [here](#).

Annual Customer Activity Growth Analysis

Customer Growth Analysis



Summary

Based on the insights from the previous graphics, we can conclude that:

- The high number of new customer growth marks the success of the marketing team in recruiting new users.
- However, it can be seen that the number of repeat customers every year was not increasing significantly since the year 2017, and decreased instead. It is important to find a solution to increase the repeat customers.

Check out the ERD [here](#).

Check out the queries [here](#).

Annual Product Category Analysis

Yearly Revenue



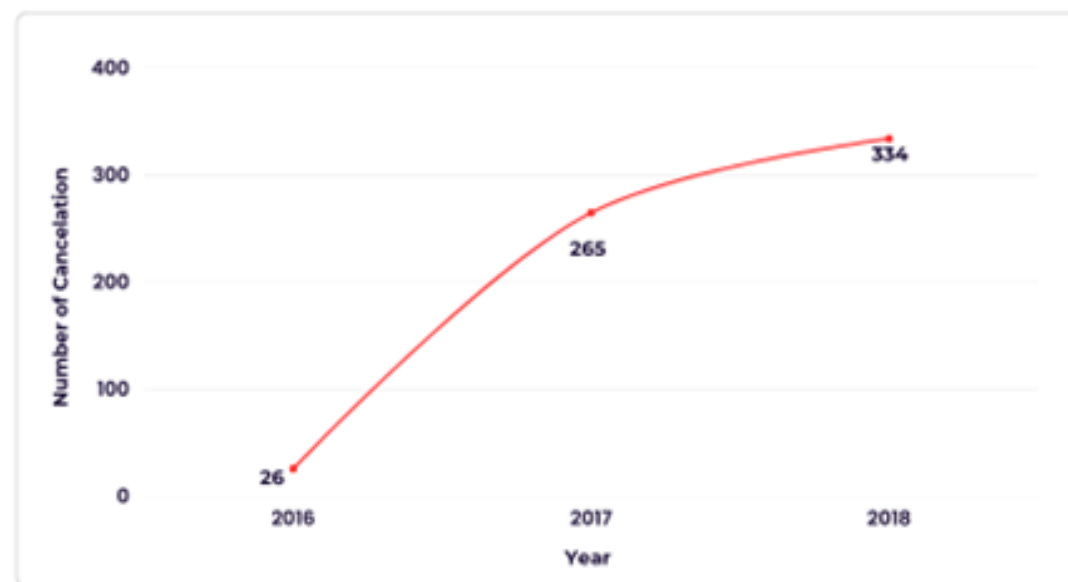
Revenue



Increase in revenue from the year 2016 to the year 2018, where:

- Significant increase in revenue of 7,085,489.22 from the year 2016 to the year 2017.
- Increase in revenue of 1,501,025.170 from the year 2017 to the year 2018.

Yearly Cancellation



Canceled Order



There is an increase in transaction cancelation from the year 2016 to the year 2018, where:

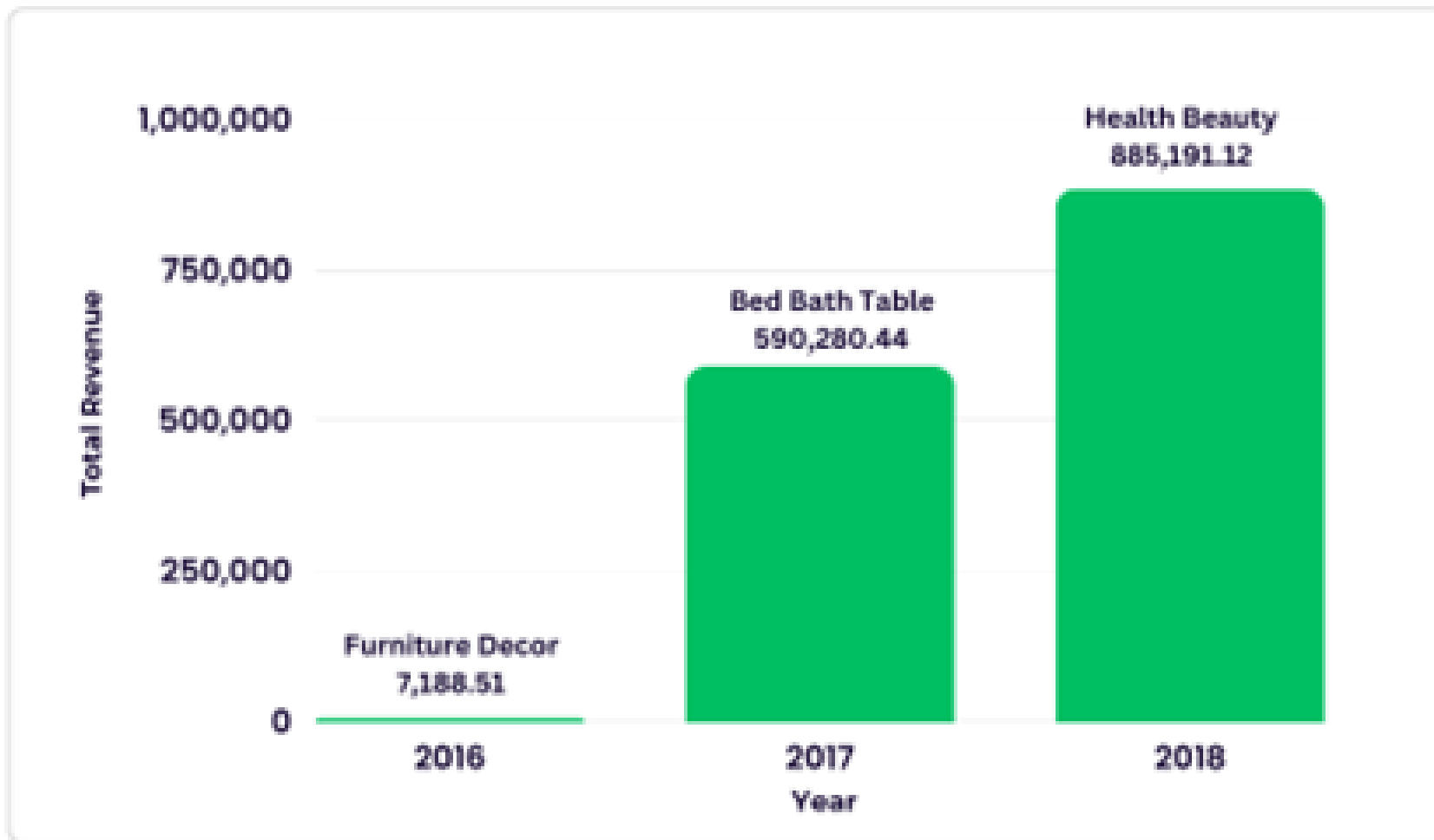
- The total of transaction cancellations increased from 26 cancellations in the year 2016 to 265 cancellations in the year 2017 (an increase of 239 cancellations).
- The total of transaction cancellations increased from 265 cancellations in the year 2017 increased to 334 cancellations in the year 2018 (an increase of 69 cancellations).

Check out the ERD [here](#).

Check out the queries [here](#).

Annual Product Category Analysis

Yearly Highest Revenue Category



Year 2016: Furniture Decor category has the highest revenue of 7,188.51



Year 2017: Bed Bath Table category has the highest revenue of 590,280.44

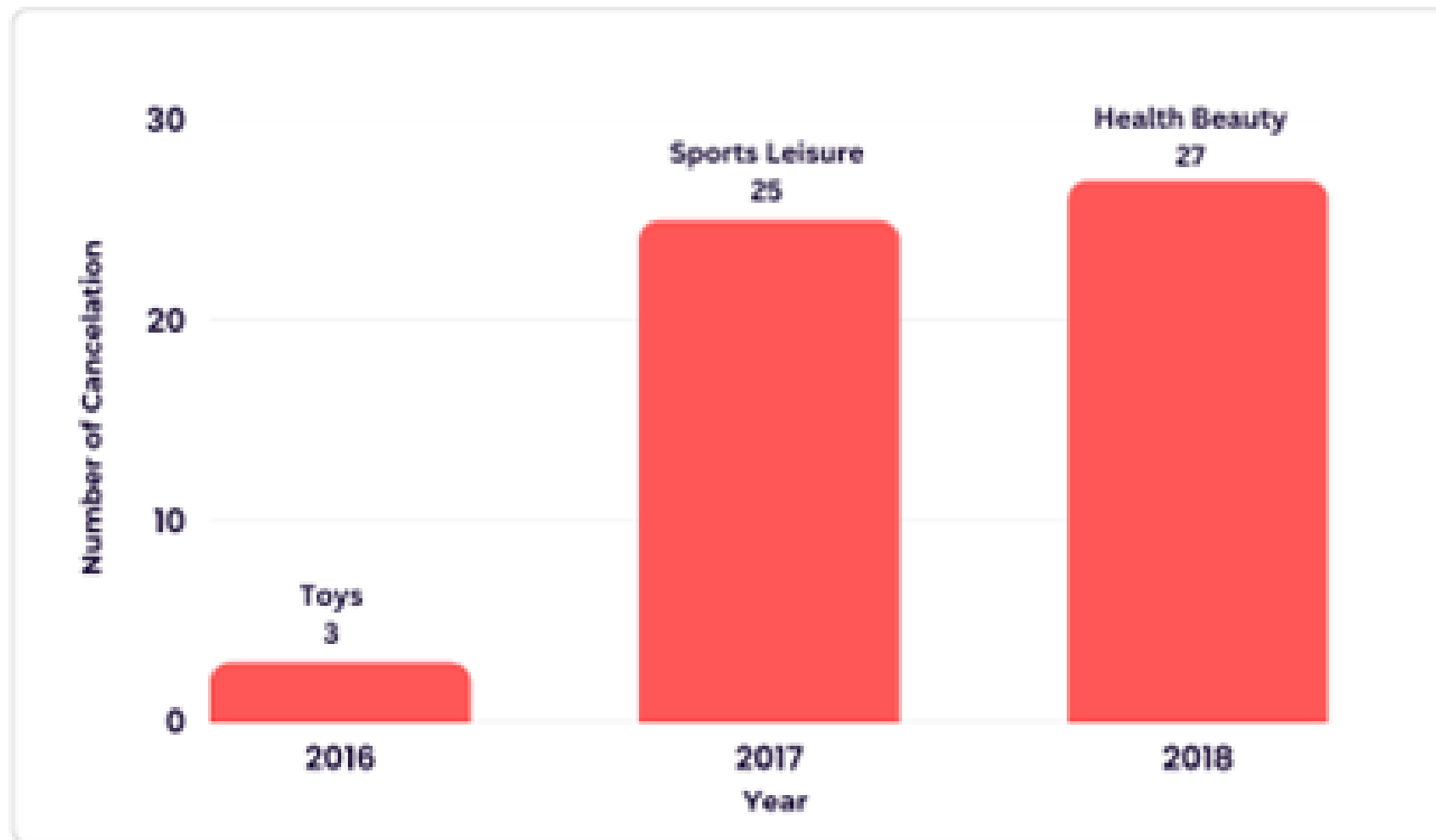


Year 2018: Health & Beauty category has the highest revenue of 885,191.12

Check out the ERD [here](#).
Check out the queries [here](#).

Annual Product Category Analysis

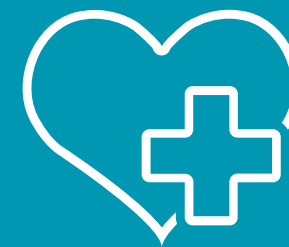
Yearly Highest Cancellation Category



Year 2016: Toys category has the highest cancellations of 3.



Year 2017: Sports Leisure category has the highest cancellation of 25.



Year 2018: Health & Beauty category has the highest cancellation of 27.

Check out the ERD [here](#).
Check out the queries [here](#).

Annual Product Category Analysis

Conclusion:

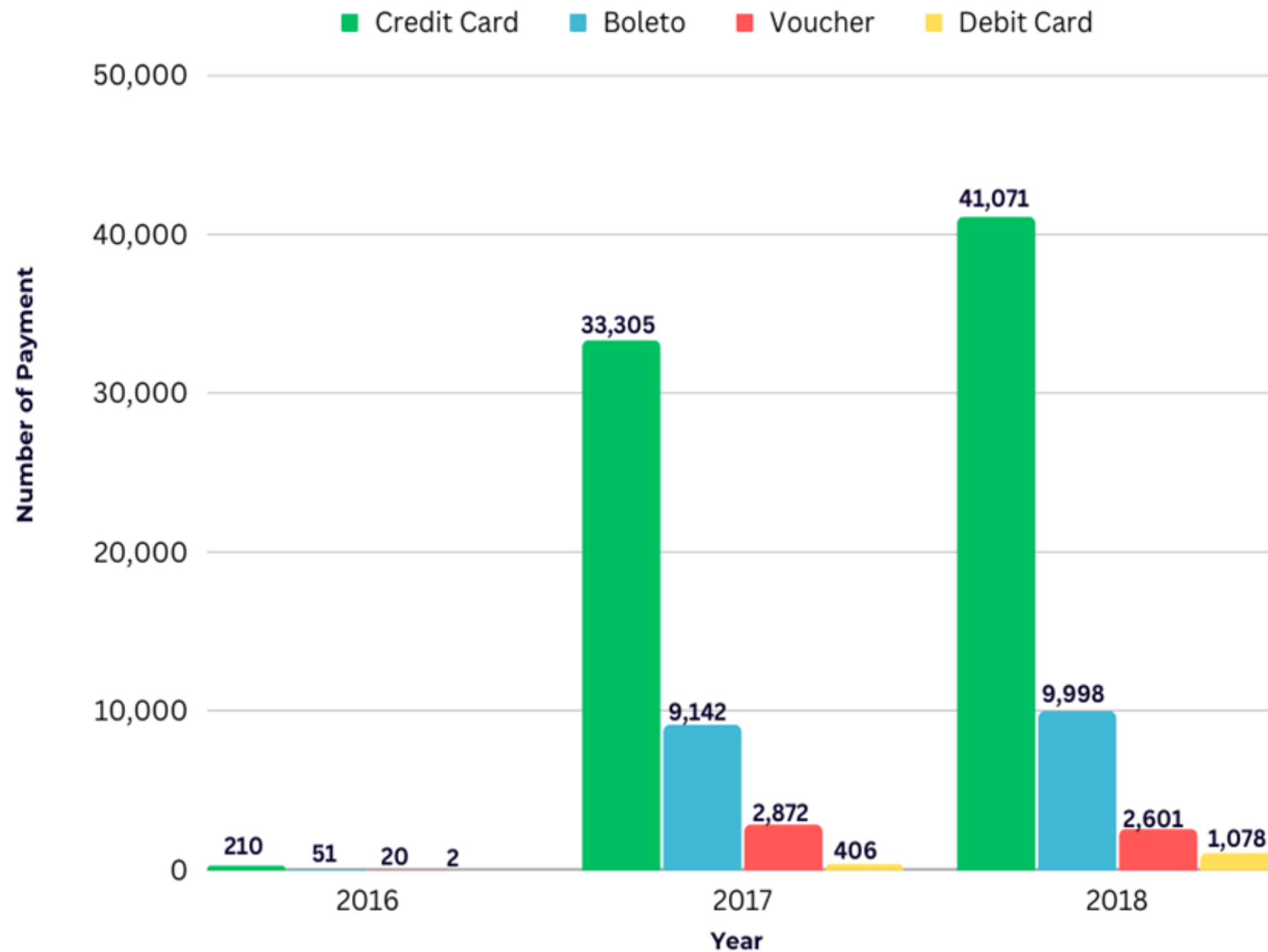
- The increase in total revenue that occurs every year from the year 2016 to the year 2017 can be expected to occur due to a large number of transactions and is **directly proportional** to the very significant growth of new customers from 2016 to 2017. The increase also occurred in 2018, but not too significant if compared to the previous increase, which is also in accordance with the not-very significant customer growth in 2018.
- The increase in cancelations that occurred in the year 2017 and year 2018 was insignificant, but it is recommended to carry out an analysis of the cancellations made and continue with finding solutions to minimize the occurrence of cancelations that can harm merchants and e-commerce.
- The difference in the categories with the highest total revenue each year can be caused by changing trends, such as certain types of products or brands that have promotions or good reviews, increased customer demand, and so on.
- The highest cancellation category which differs every year can also be influenced by various factors, such as previous bad customer reviews, product originality, licenses, environmental sustainability issues, or black campaigns that occur against certain brands.

Check out the ERD [here](#).

Check out the queries [here](#).

Analysis of Annual Payment Type Usage

Yearly Trend of Payment Type



Based on the graph, there are 4 types of payments consisting of:

- Credit card.
- Boleto (Popular payment type used in Brazil, South America. [See here](#)).
- Vouchers.
- Debit Cards.

Consistently from the year 2016 to the year 2018, the trend of the most frequently used payment types is as follows:

- Credit card.
- Boleto.
- Vouchers.
- Debit Cards.

Check out the ERD [here](#).

Check out the queries [here](#).

Analysis of Annual Payment Type Usage

Conclusion:

- The increasing number of payments from the year 2016 to the year 2018 is directly proportional to the growth of customers and the number of occurring transactions.
- Credit Card is the most popular payment method from the year 2016 to the year 2018, with the usage percentage reaching 74% in the year 2016, 73% in the year 2017, and 75% in the year 2018. The high use of Credit cards can be influenced by several factors, such as the availability of payment in installments that might be preferred by certain customers, the benefits of a good credit score, or points awarded for each transaction paid using a credit card.
- Boleto is the second most popular payment method besides Credit Cards. Boleto is an official payment method in Brazil, with usage rates of 18% in the year 2016, 20% in the year 2017, and 18% in the year 2018.
- Vouchers and Credit Cards are in third and fourth place from the year 2016 to the year 2018. Based on this insight, it can be concluded that payments using Debit Cards are not very popular. As for vouchers, there is a possibility that vouchers are only available in limited quantities, or only given to certain customer segments (can be further analyzed), so not many customers can use these vouchers.

Check out the ERD [here](#).

Check out the queries [here](#).

Thank You



devinalim5@gmail.com



[Devina Lim](#)

