



Created by:

Muhammad Miftah Thaha

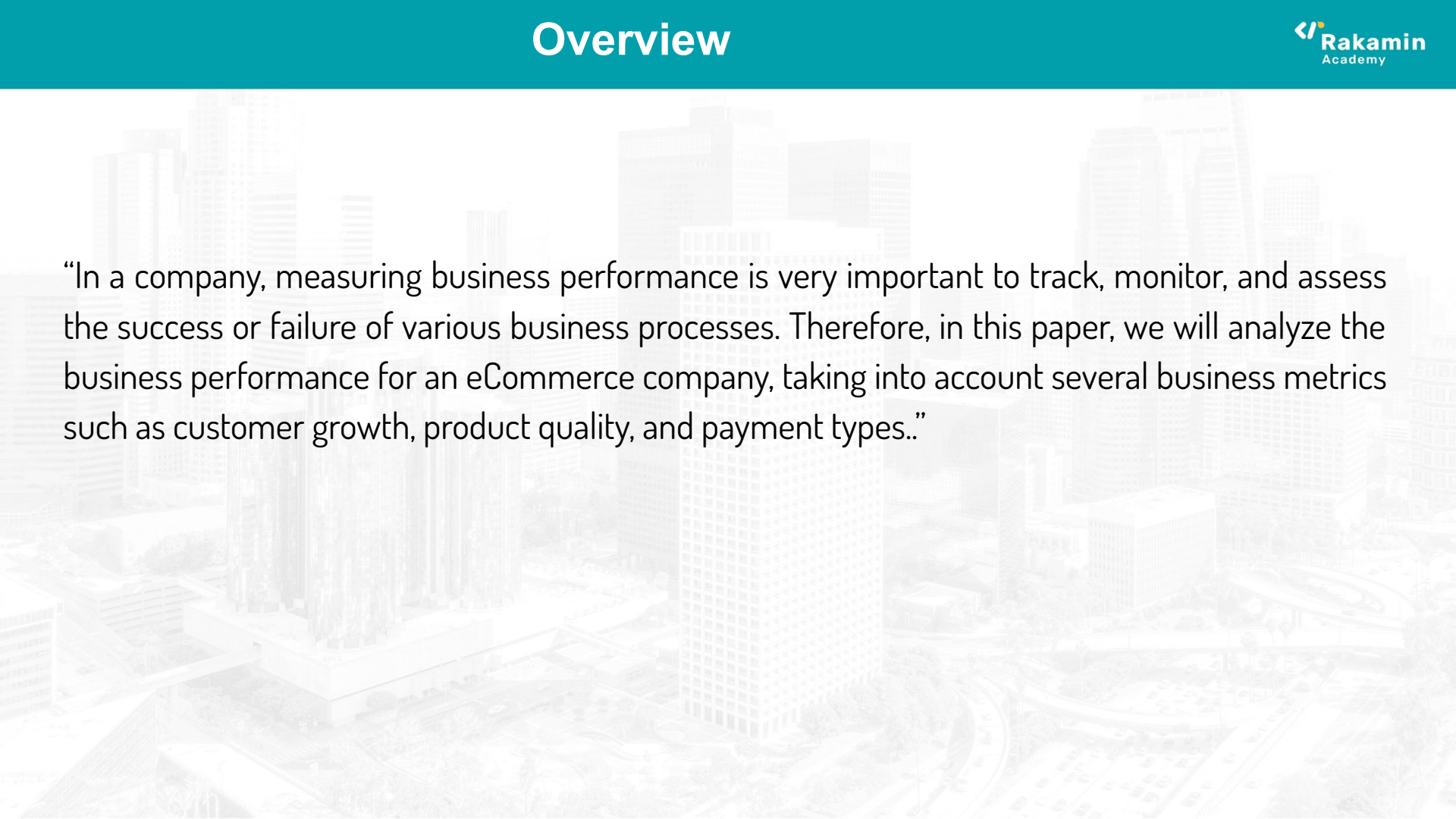
muhammadmiftaht@gmail.com

<https://www.linkedin.com/in/miftahthaha/>

<https://github.com/miftahthaha>

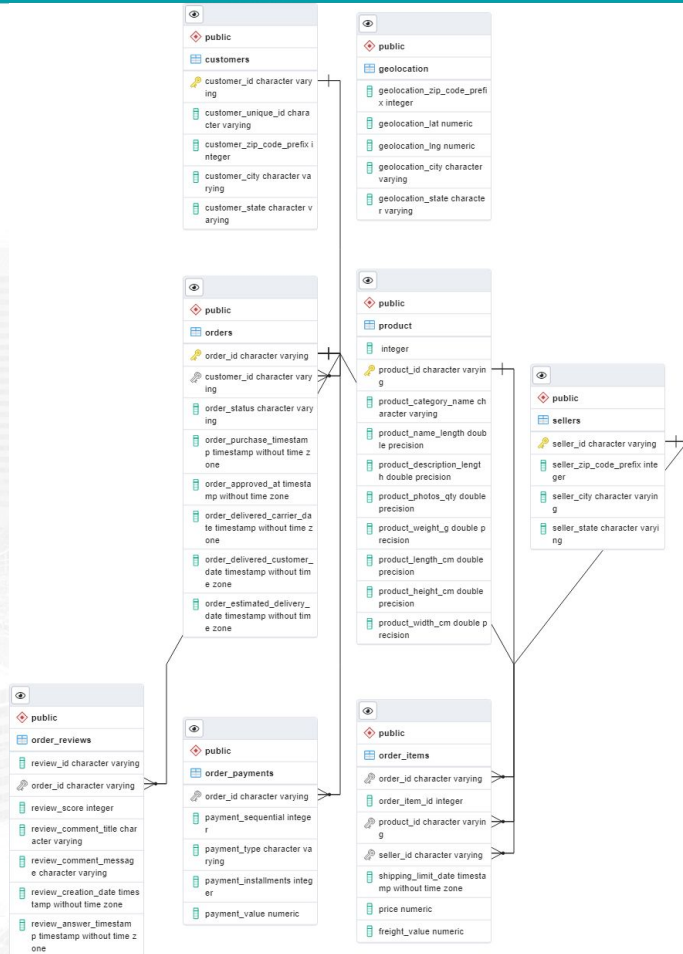
Analyzing eCommerce Business Performance with SQL

Miftah is a recent graduate with a Bachelor's degree Electrical Engineer who have interest in Data Analytics and Science and have a strong foundation in statistical modeling, data analysis, and programming. As a Junior Data Scientist, he has experience through his final project in building and implementing machine learning models, analyzing complex data sets, and creating visualizations to communicate insights. He is a fast learner with excellent problem-solving skills and a passion for using data to drive business decisions. In addition, he possess strong communication and collaboration skills, having worked on multiple team projects during his studies. With a drive to excel in his field, Miftah is seeking an opportunity to contribute his skills and knowledge to a dynamic and innovative organization as a Junior Data Scientist.

A faded, light grey background image of a city skyline with various skyscrapers and buildings.

“In a company, measuring business performance is very important to track, monitor, and assess the success or failure of various business processes. Therefore, in this paper, we will analyze the business performance for an eCommerce company, taking into account several business metrics such as customer growth, product quality, and payment types..”

Data Preparation



- Created E-Commerce Business Performance database in PostgreSQL
- Created 9 New Table (customer, geolocation, order_item, order_payment, order_review, orders, product, and seller) with respective datatypes for each values with 'CREATE TABLE' statement.
- Imported CSV data into database using 'COPY FROM' statement.
- Determine the primary and foreign key for relation in each data and create The ERD as as a result of importing the table into the database.

Annual Customer Activity Growth Analysis

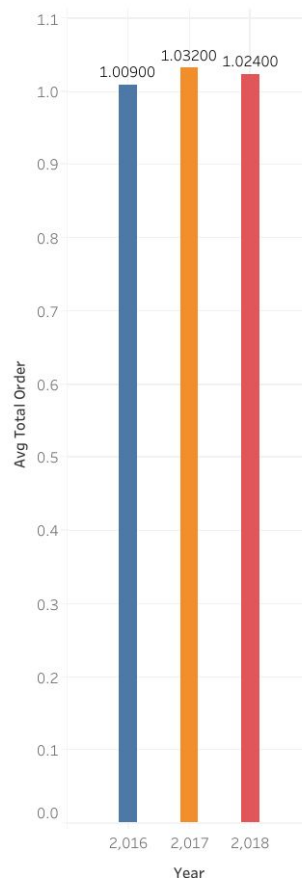
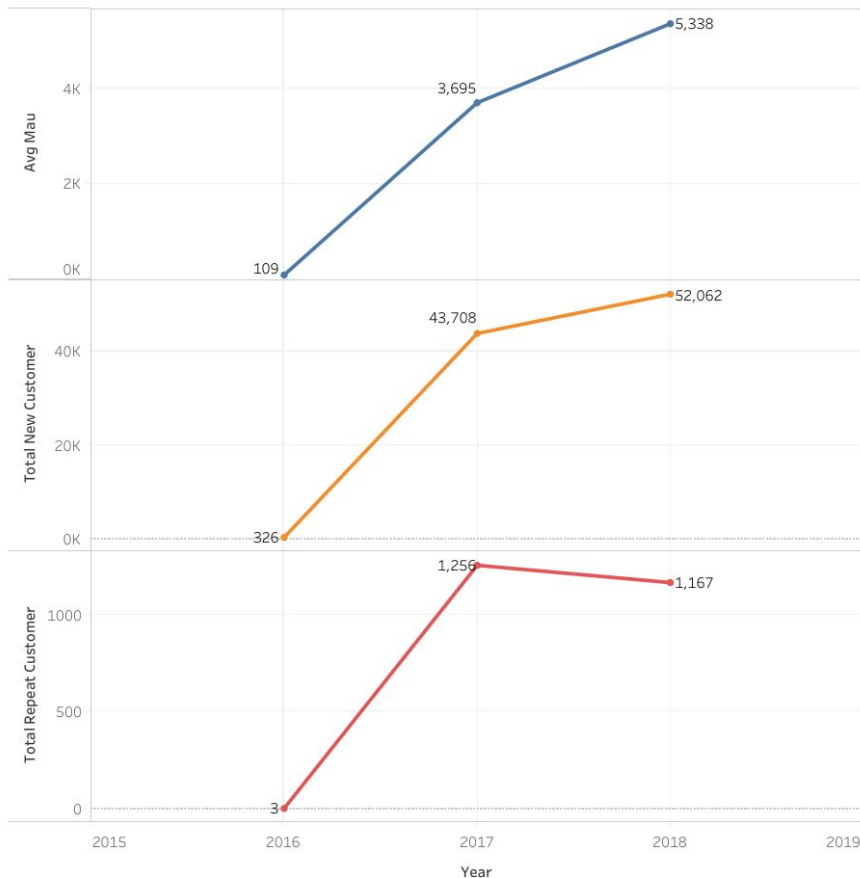
	year double precision	avg_mau numeric	total_new_customer bigint	total_repeat_customer bigint	avg_total_order numeric
1	2016	108.667	326	3	1.009
2	2017	3694.833	43708	1256	1.032
3	2018	5338.200	52062	1167	1.024

Table from querying using customer and order dataset.

- Avg_mau = Average value of Monthly Active User per year
- Total_new_customer = Total new customer per year
- Total_repeat_customer = Total customer who re-ordering per year
- Avg_total_order = Average frequency order value per year

Annual Customer Activity Growth Analysis

Annual Customer Activity Growth Analysis



The above result shows the average monthly active users (MAU), the number of new customers, the number of repeat customers, and the average frequency of orders for each year from 2016 to 2018.

In 2016, the company had an average of 108.667 monthly active users with a very low number of repeat customers and a slightly above-average frequency of orders at 1.009.

For more details, you can see all file [here](#)

Annual Customer Activity Growth Analysis

This indicates that the company was still in its early stages in 2016 and was primarily focused on acquiring new customers.

In 2017, the company's average monthly active users increased significantly to 3,694.833 with a significant increase in both new and repeat customers. The average frequency of orders also increased slightly to 1.032. This indicates that the company was successful in acquiring and retaining customers in 2017.

In 2018, the company continued to grow with an average of 5,338.200 monthly active users, with a similar number of new customers as in 2017 and a slightly lower number of repeat customers. The average frequency of orders decreased slightly to 1.024, indicating that customers were not ordering as frequently as they did in 2017. Overall, the company's performance was still strong in 2018, but it may need to focus on increasing customer retention and order frequency to maintain growth in the future.

Annual Product Category Quality Analysis

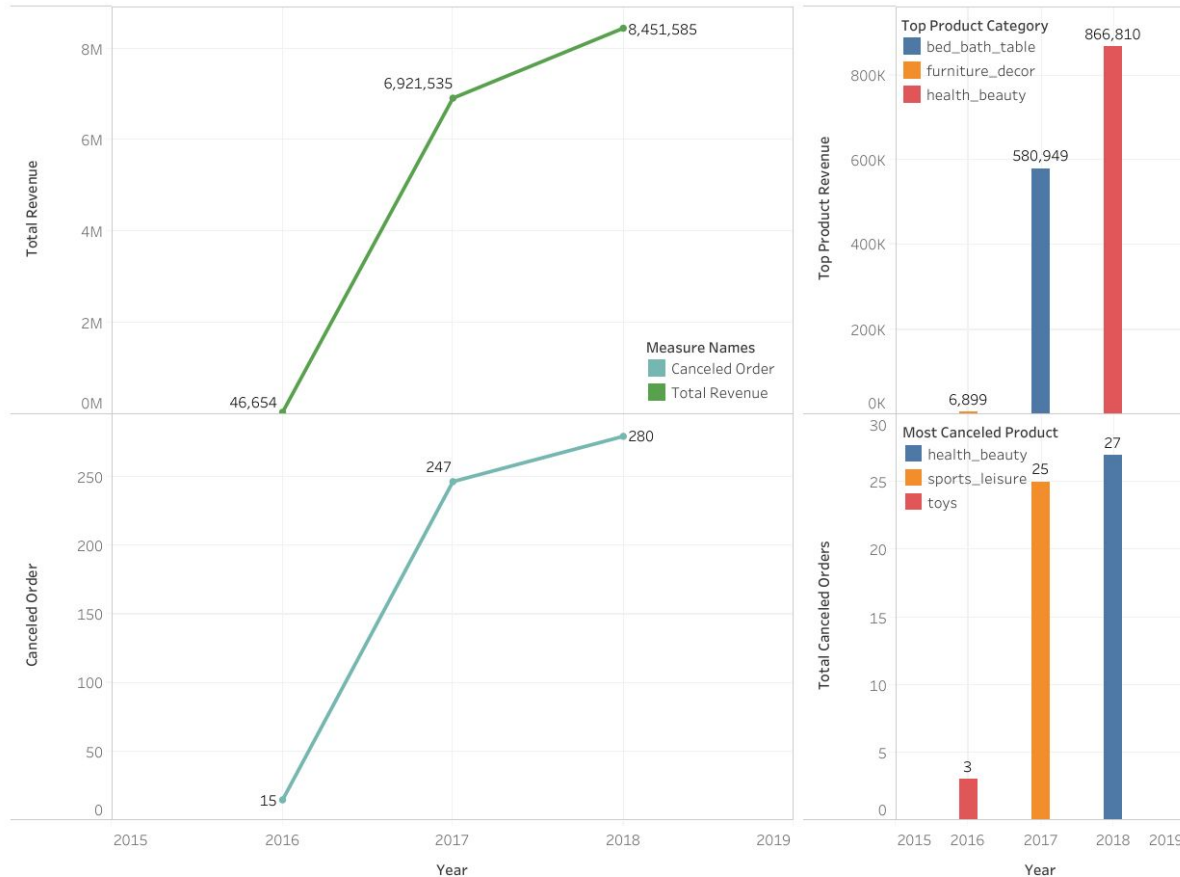
	year double precision 🔒	total_revenue numeric 🔒	canceled_order numeric 🔒	top_product_category character varying 🔒	top_product_revenue numeric 🔒	most_canceled_product character varying 🔒	total_most_canceled_orders numeric 🔒
1	2016	46654	15	furniture_decor	6899	toys	3
2	2017	6921535	247	bed_bath_table	580949	sports_leisure	25
3	2018	8451585	280	health_beauty	866810	health_beauty	27

Table from querying using order_items and orders dataset.

- total_revenue = Total company revenue per year (price + freight value)
- canceled_order = Total company canceled order per year
- top_product_category = Rank 1 sold product category per year
- top_product_revenue = Rank 1 Revenue for sold product per year
- most_canceled_product = Most canceled product per year
- total_most_canceled_order = Most canceled order per year

Annual Product Category Quality Analysis

Annual Product Category Quality Analysis



Based on the visualization results, we can interpret the performance of the eCommerce business over the years as follows:

In 2016, the business had a total revenue of 46,654 with only 15 canceled orders. The top product category was furniture_decor with the highest revenue generated by the top product being 6,899.

For more details, you can see all file [here](#)

Annual Product Category Quality Analysis

However, the most canceled product category was toys, with a total of 3 canceled orders.

In 2017, the business had a significant increase in revenue, totaling 6,921,535, with 247 canceled orders. The top product category was bed_bath_table, and the highest revenue was generated by the top product in that category, which was 580,949. However, the most canceled product category was sports_leisure, with a total of 25 canceled orders.

In 2018, the business had a further increase in revenue, totaling 8,451,585, with 280 canceled orders. The top product category was health_beauty, and the highest revenue was generated by the top product in that category, which was 866,810. Similar to the previous year, the most canceled product category was health_beauty, with a total of 27 canceled orders.

Annual Product Category Quality Analysis

Overall, we can see that the eCommerce business has been growing over the years, with an increase in revenue and a relatively low number of canceled orders. The top product categories and highest revenue generating products vary between years, but the most canceled product category remained the same with `top_product_category` in the last year. This could suggest that there were different issues with the products or services provided in each year that led to customers canceling their orders. The business may benefit from further analysis of the reasons behind canceled orders, especially in the `health_beauty` category, to address and work to improve the customer experience to reduce these cancellations.

Analysis of Annual Payment Type Usage

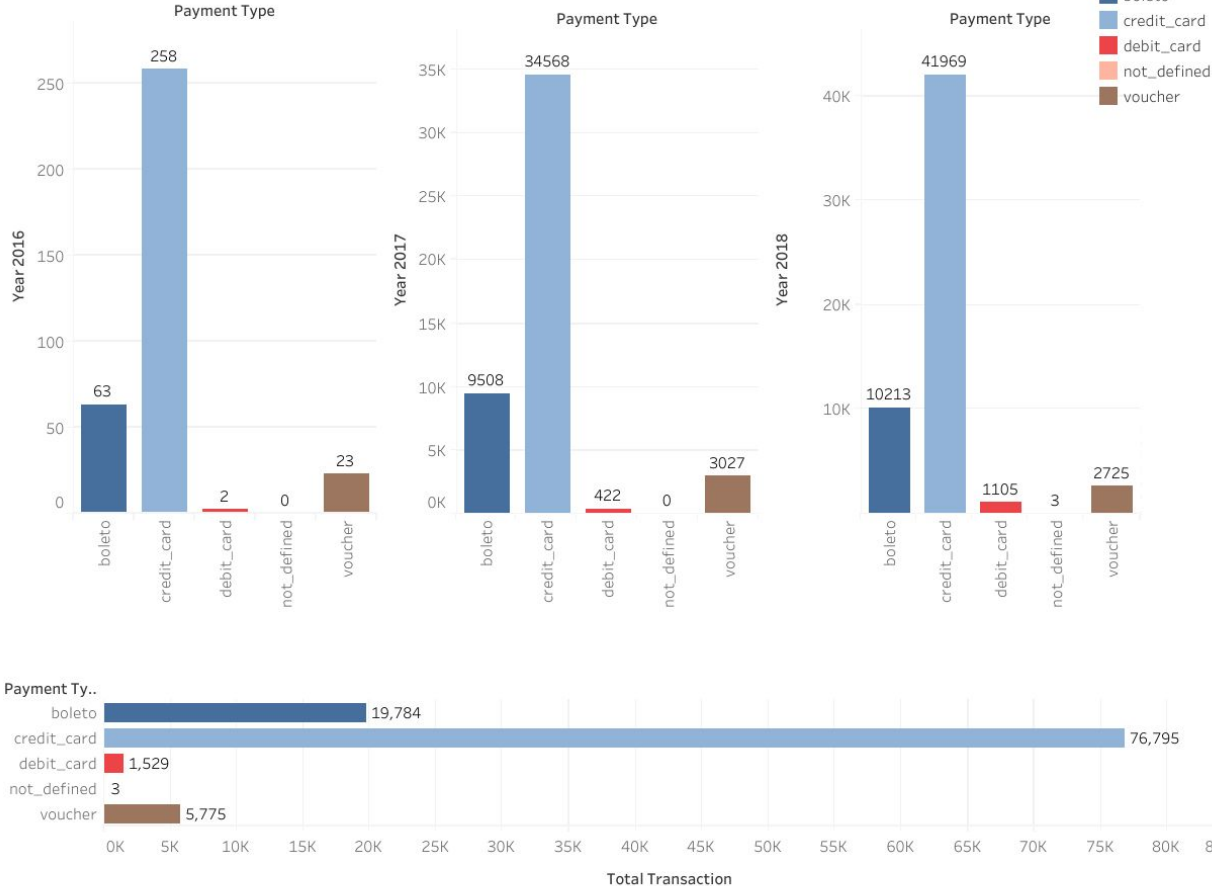
	payment_type character varying 🔒	total_transaction bigint 🔒	year_2016 bigint 🔒	year_2017 bigint 🔒	year_2018 bigint 🔒
1	credit_card	76795	258	34568	41969
2	boleto	19784	63	9508	10213
3	voucher	5775	23	3027	2725
4	debit_card	1529	2	422	1105
5	not_defined	3	0	0	3

Table from querying using order_payments and orders dataset.

- payment_type = The method used by customers to make payment for their purchases
- total_transaction = Total number of transactions made by customers in a given period of time
- year_2016 = Total transactions in specific year
- year_2017 = Total transactions in specific year
- year_2018 = Total transactions in specific year

Analysis of Annual Payment Type Usage

Analysis of Annual Payment Type Usage



The visualization beside provides information about the distribution of payment types used by customers to make transactions in three consecutive years: 2016, 2017, and 2018. The payment type "credit_card" is the most commonly used payment method for transactions, with a total of 76,795 transactions over the three-year period.

For more details, you can see all file [here](#)

"Boleto" is the second most used payment type, with a total of 19,784 transactions. "Voucher" and "debit_card" are used less frequently than the other payment types, with 5,775 and 1,529 total transactions, respectively. There are also a small number of transactions (3) where the payment type is not defined.

It's also interesting to note that the number of transactions made with each payment type differs across the years. For example, the number of transactions made using "credit_card" increased greatly each year, while the number of transactions made using "boleto" increased slightly in 2018 and debit_card remained relatively consistent. Meanwhile, the number of transactions made using "voucher" increased in 2017 compared to 2016, but decreased in 2018.

In other hand, it would be a good practice to investigate the "not_defined" payment type further to determine the reason for its occurrence. It could be a data entry error or an actual payment type used by a small number of customers. Understanding the nature of this payment type could provide insights into the data and help improve the accuracy of future analyses.

Thank You!

Supported by:
Rakamin Academy
Career Acceleration School
www.rakamin.com