Analyzing eCommerce Business Performance with SQL

Created by:
Moh. Harwin Prayoga
mh.prayoga11@gmail.com
https://www.linkedin.com/in/mhprayoga/

"An enthusiastic learner, analytical, and flexible graduate of bachelor's degree of Engineering Physics at Institut Teknologi Sepuluh Nopember. I had experience in leadership and teamwork in various organizations and events. Moreover, I have a decent ability in English and operating various data programming software such as MS Excel, Python, SQL, etc. I am excited about seeking a challenge in the field of data where my passion, education, and training background can be fully utilized. "

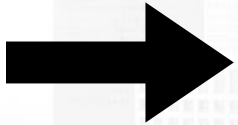
Supported by:
Rakamin Academy
Career Acceleration School
www.rakamin.com

Overview

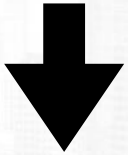
"In a company, measuring business performance is very important to track, monitor and assess the success or failure of various business processes. Therefore, in this paper we will analyze business performance for an eCommerce company, taking into account several business metrics, namely customer growth,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

Data Preparation – Task 1

- There are 8 datasets to be used in this project, i.e.:
  - **customers**
  - **geolocation**
  - **order_items**
  - **order_payments**
  - **order_reviews**
  - **orders**
  - **product**
  - **sellers**



- Cleaning geolocation dataset due invalid values. [Code]

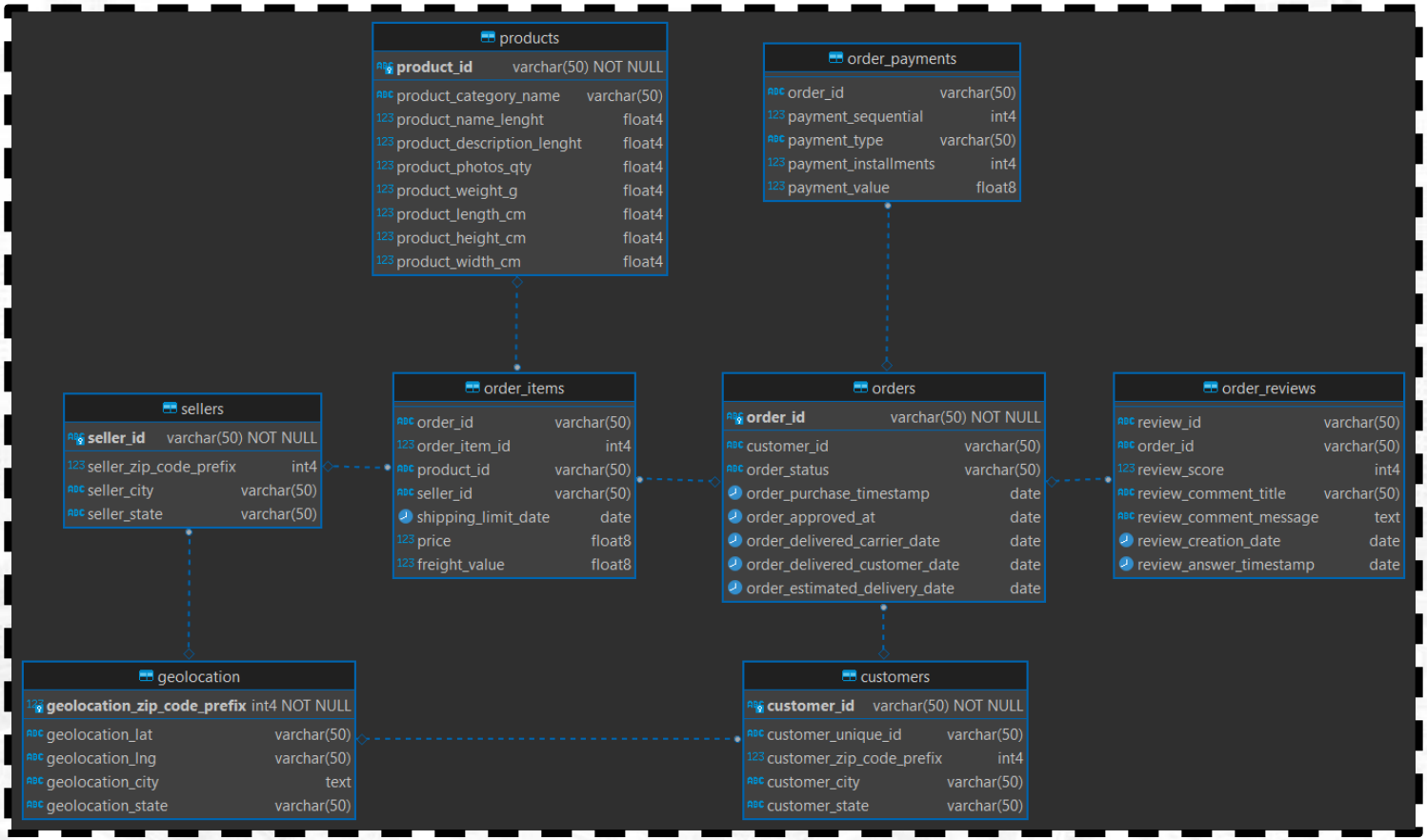


- Choosing primary key and foreign key of each dataset according to Data Relationship.

- Creating Entity Relationship Diagram (ERD) with DBeaver.

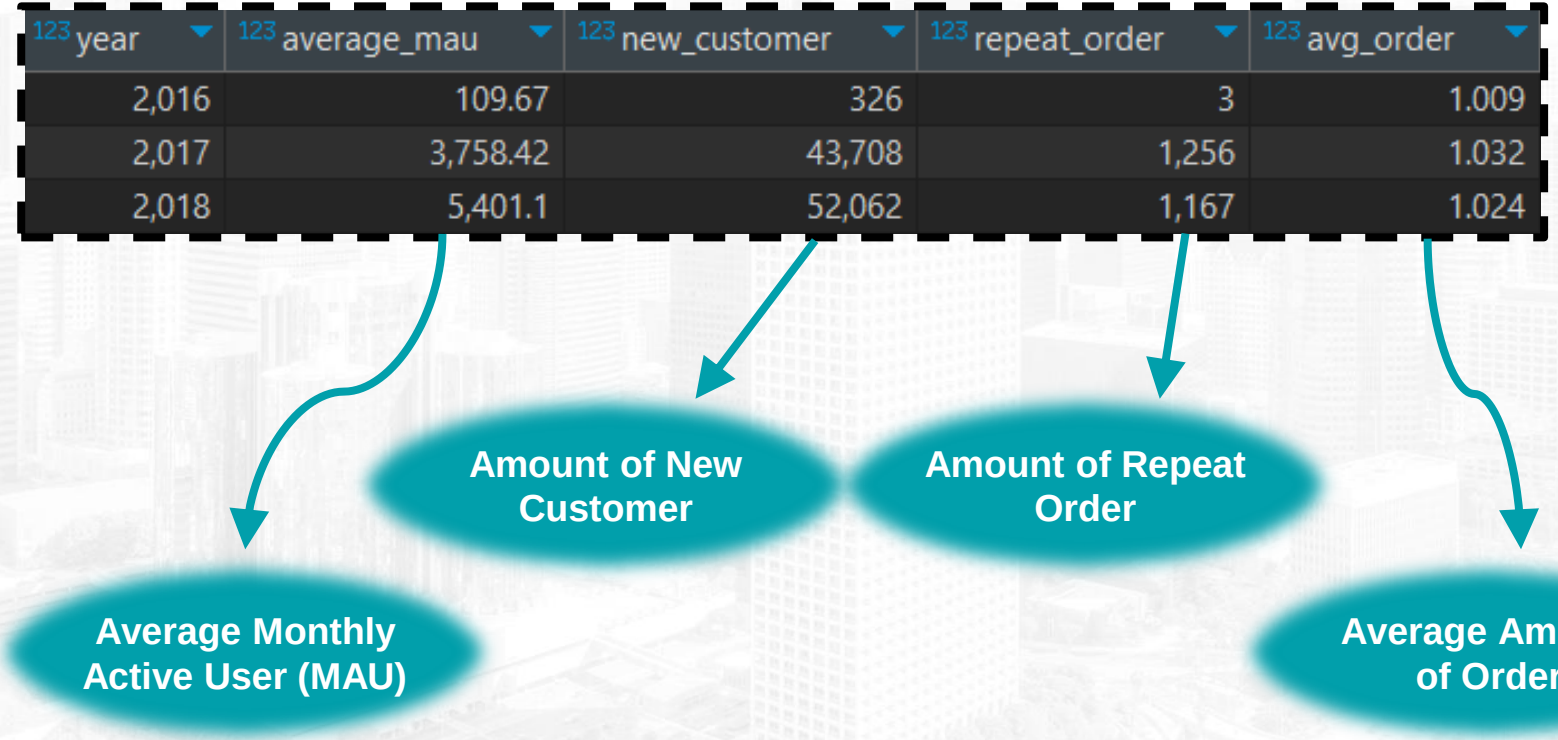
Entity Relationship Diagram (ERD) – Task 1

Rakamin Academy



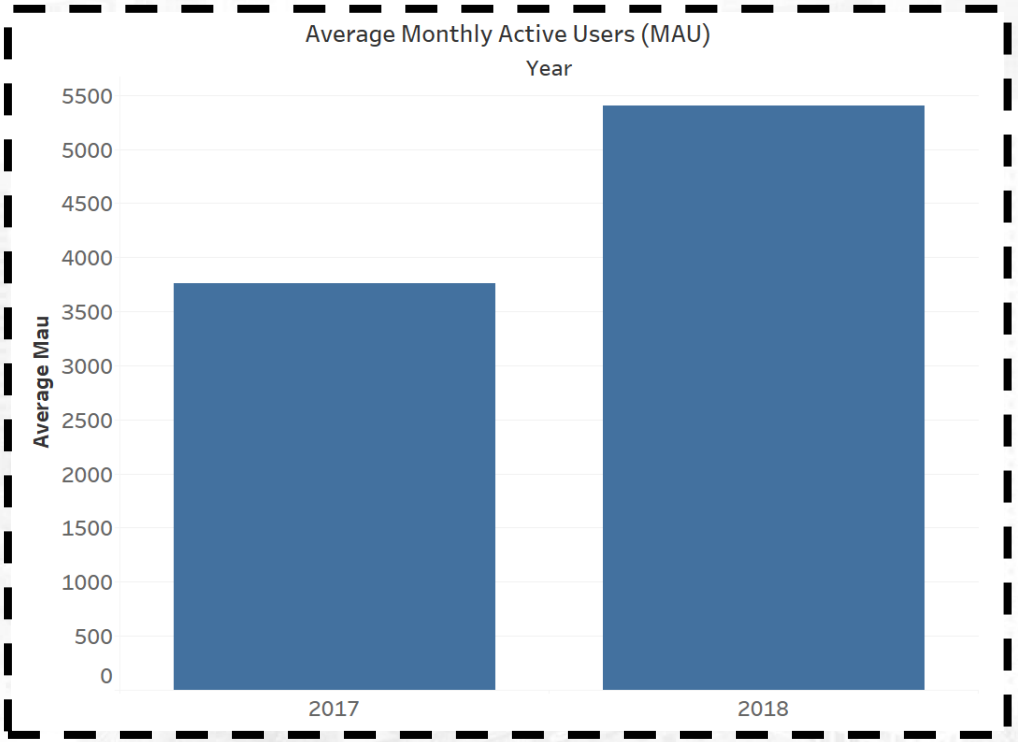
Exploratory Data Analysis (EDA) – Task 2

year	average_mau	new_customer	repeat_order	avg_order
2,016	109.67	326	3	1.009
2,017	3,758.42	43,708	1,256	1.032
2,018	5,401.1	52,062	1,167	1.024



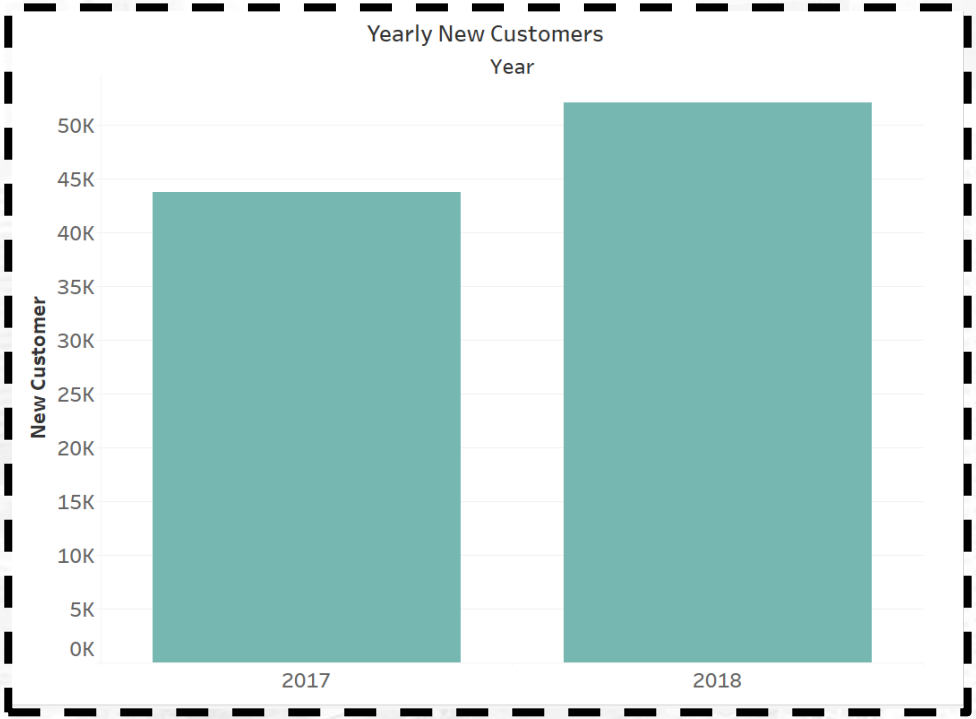
[Task 2 source code]

Exploratory Data Analysis (EDA) – Task 2



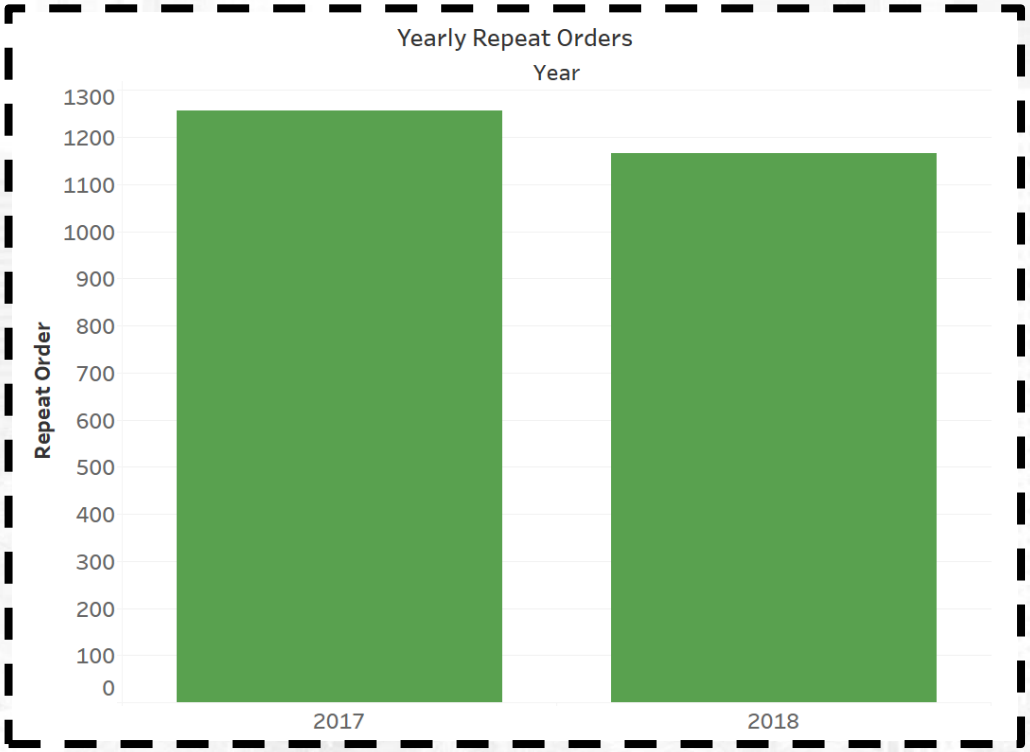
The dataset contains transaction data from September 2016 to December 2018. Because 2016 has only 4 months of data, the year 2016 has a highly different amount of data compared with other years. The chart on the left shows that **Monthly Active Users (MAU) is significantly increasing.**

Exploratory Data Analysis (EDA) – Task 2



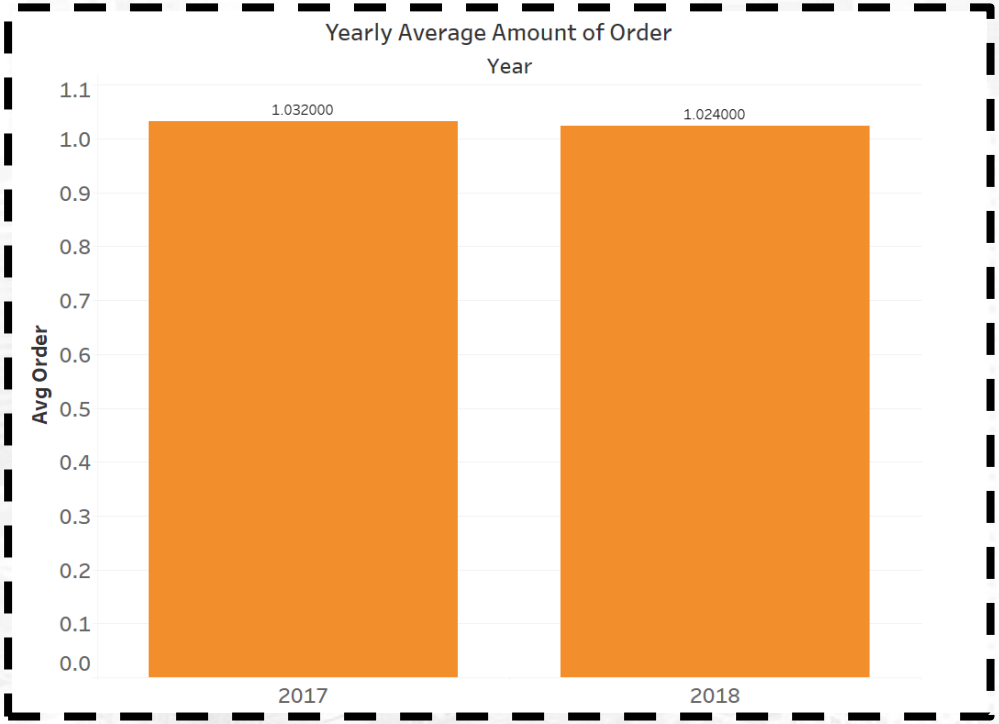
Following with increment of MAU, the Amount of new customers also **increased in 2017 to 2018.**

Exploratory Data Analysis (EDA) – Task 2



Although we have MAU and New Customers increased, the amount of customers that did repeat order are **decreased a little in 2017 to 2018.**

Exploratory Data Analysis (EDA) – Task 2



The decrement in number also happened with Yearly Repeat Orders, the Average Amount of Order **decreased in 2017 to 2018** which means most of them (customers) still ordered only 1 item.

Exploratory Data Analysis (EDA) – Task 3

Rakamin Academy

Top Gained Revenue Category

Amount of Cancel

Amount of Top Cancel Category

	year	yearly_revenue	top_revenue_category	top_category_revenue	mnt_yearly_cancel	top_cancel_category	category_mnt_cancel
1	2,016	46,653.74	furniture_decor	7,188.51	26	toys	3
2	2,017	6,921,535.239999724	bed_bath_table	590,280.4400000052	265	sports_leisure	25
3	2,018	8,451,584.769999897	health_beauty	885,191.1199999922	334	health_beauty	27

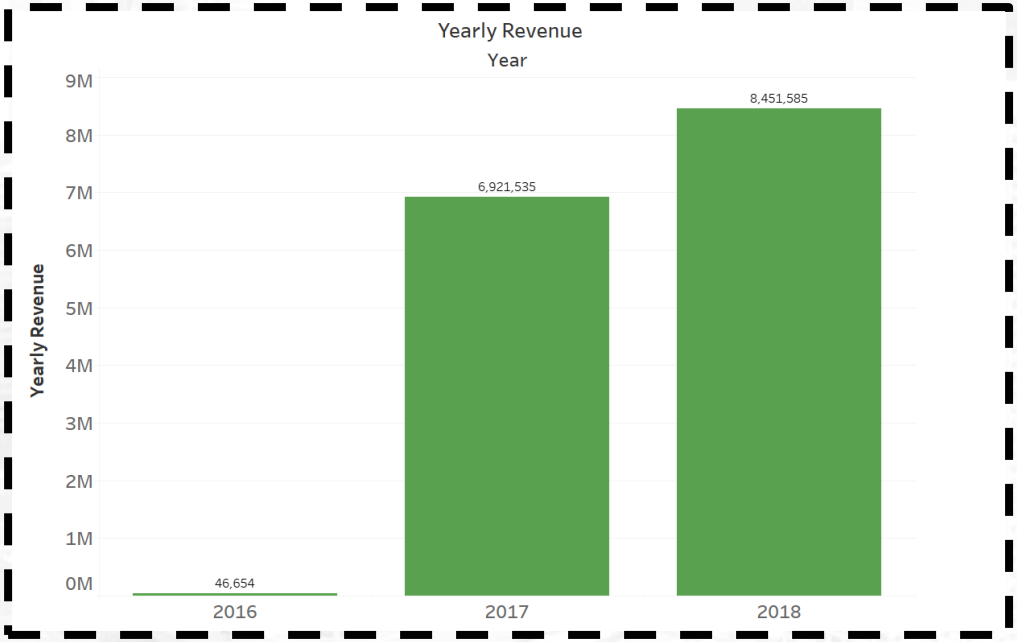
Average Revenue

Revenue of Top Revenue Category

Category of Top Amount of Cancel

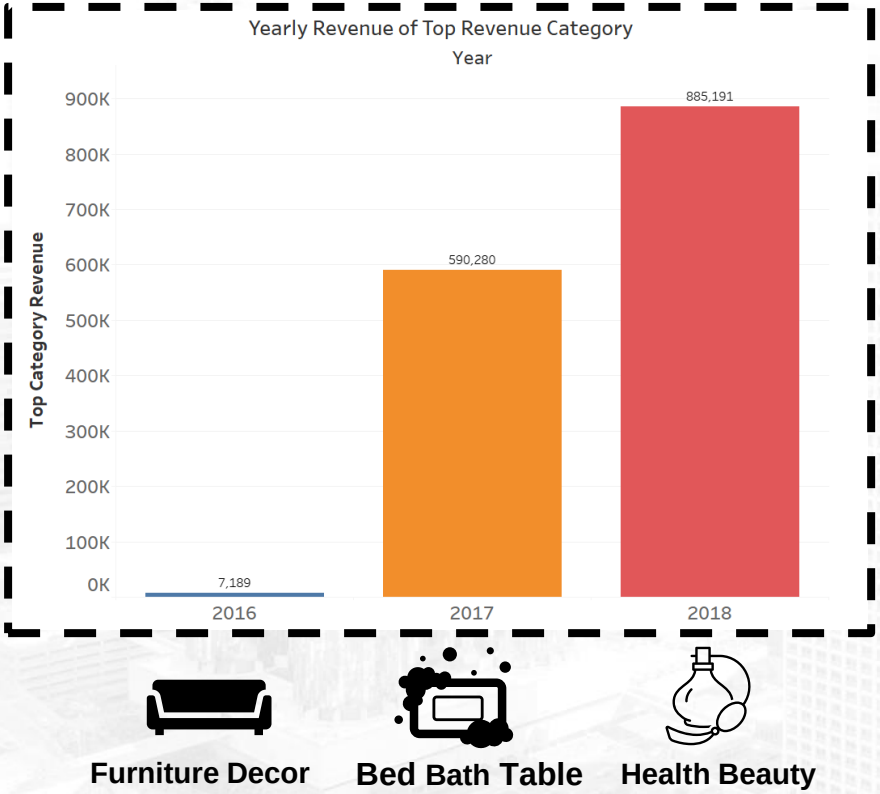
[Task 3 source code]

Exploratory Data Analysis (EDA) – Task 3



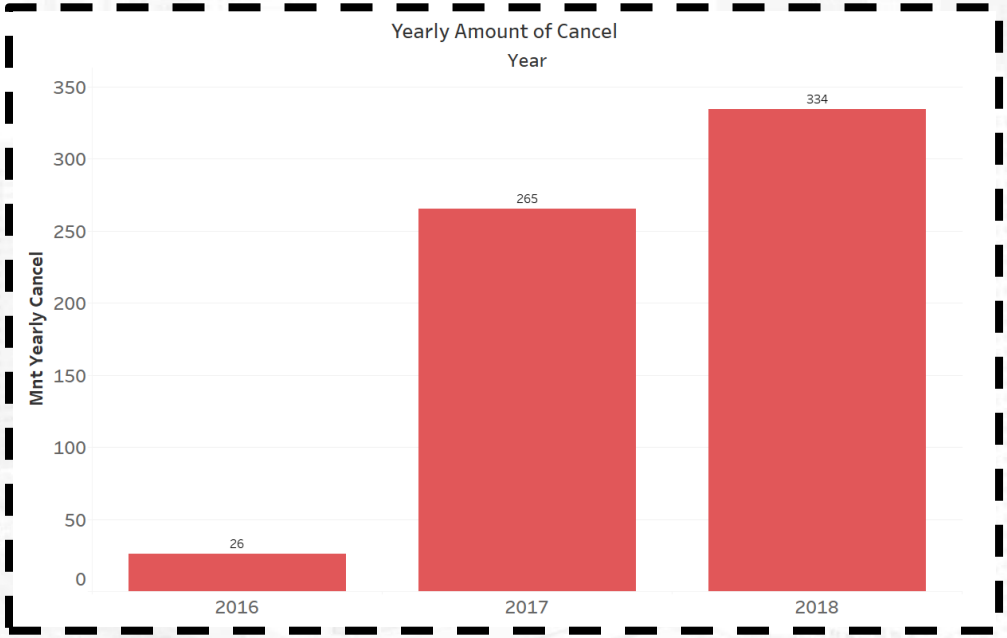
In this analysis, we can see that the revenue is **increasing each year** along with increment of MAU and New Customers**.**

Exploratory Data Analysis (EDA) – Task 3



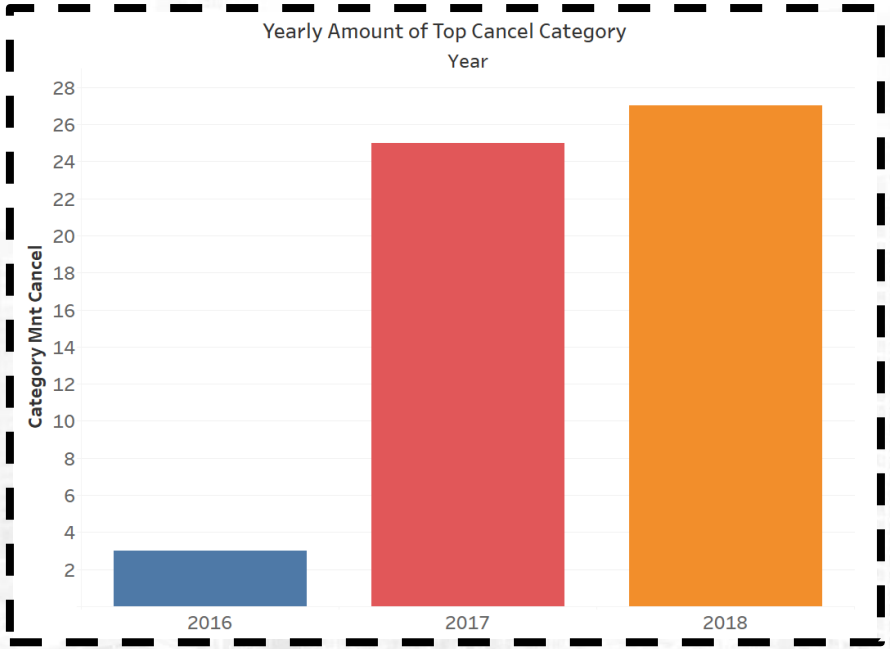
From the chart on the left, the top revenue category in 2016 is **Furniture Décor**, 2017 is **Bed Bath Table**, 2018 is **Health Beauty**

Exploratory Data Analysis (EDA) – Task 3



Although the average of orders is decreasing each year, the **yearly amount of cancel is increased** which surely did affect to average of orders per customer.

Exploratory Data Analysis (EDA) – Task 3



Toys

Sport Leisure

Health Beauty

From the chart on the left, the category that has the highest number of cancellation in 2016 is **Toys**, 2017 is **Sport Leisure**, and 2018 is **Health Beauty**.

What is interesting is, that the **Health Beauty category was also the top revenue category in 2018** which means the Health Beauty has been the market trend of the year 2018.

Exploratory Data Analysis (EDA) – Task 4

Rakamin Academy

Amount of Payments by Year

	payment_type	year_2016	year_2017	year_2018	pct_change_2017_2018
1	not_defined	0	0	3	[NULL]
2	debit_card	2	422	1,105	1.62
3	credit_card	258	34,568	41,969	0.21
4	boleto	63	9,508	10,213	0.07
5	voucher	23	3,027	2,725	-0.1

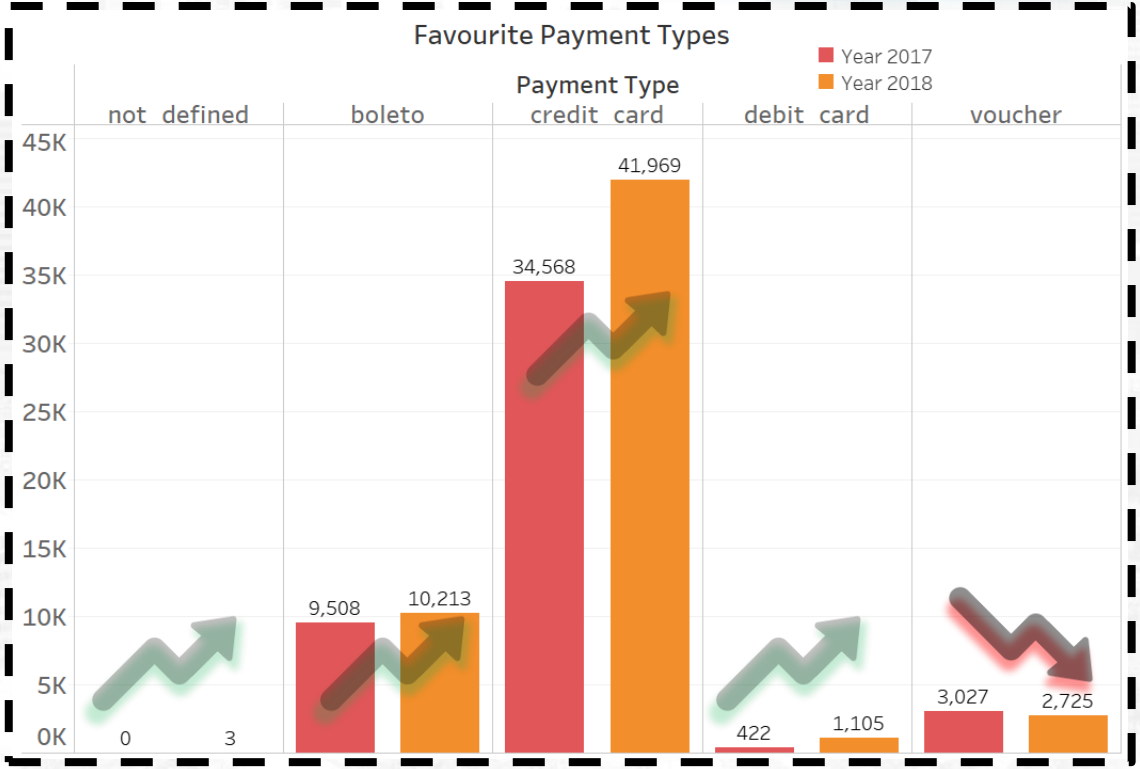
Amount of Payments by All-Time

	payment_type	amount
1	not_defined	3
2	boleto	19,784
3	debit_card	1,529
4	voucher	5,775
5	credit_card	76,795

[Task 4 source code]

Exploratory Data Analysis (EDA) – Task 4

Favourite Payment Types



The chart on the left shows that **Credit Card is the most used payment type** with a significant difference from other payment types.

On the other hand, usage of **voucher has slightly decreased than the previous year**.

This occurred probably due to cooperation with credit card partners that offered more interesting benefits to the customers such as discounts, cashback, or membership points.

Feel free to reach me at:

LinkedIn :
https://www.linkedin.com/in/mhprayoga/

GitHub:
https://github.com/mpythree

Medium:
https://medium.com/@muhammadharwinprayoga

This Portfolio:
https://github.com/mpythree/analyzing_business_performance_sql