



# *Technical Test* *@ Perqara*

Muhammad Rivan Hasri

# *Table of Contents*

**Section 1** : Project background

**Section 2** : Exploratory Data Analysis

**Section 3** : Clustering (RFM Segmentation)

**Section 4** : Business insight & Recommendation



# *Section 1*

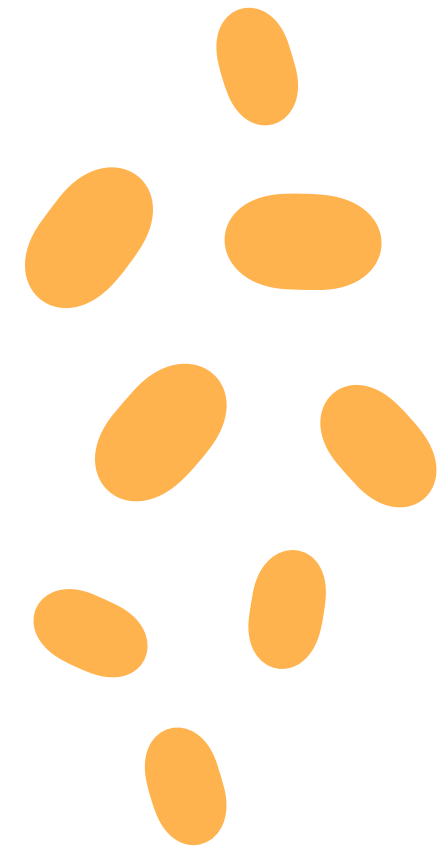
## *Project Background*



# *Background Problem*

In today's competitive market, understanding your customer base and leveraging data is key to unlocking sustainable growth. For this store to truly thrive and reach its full potential, it's essential to go beyond surface-level metrics and delve into the heart of customer behavior.

This analysis will utilize **data analysis** and **customer segmentation** to provide this store with a roadmap for strategic decision-making and enhanced customer experiences.





# *Section 2*


## *Exploratory Data Analysis*



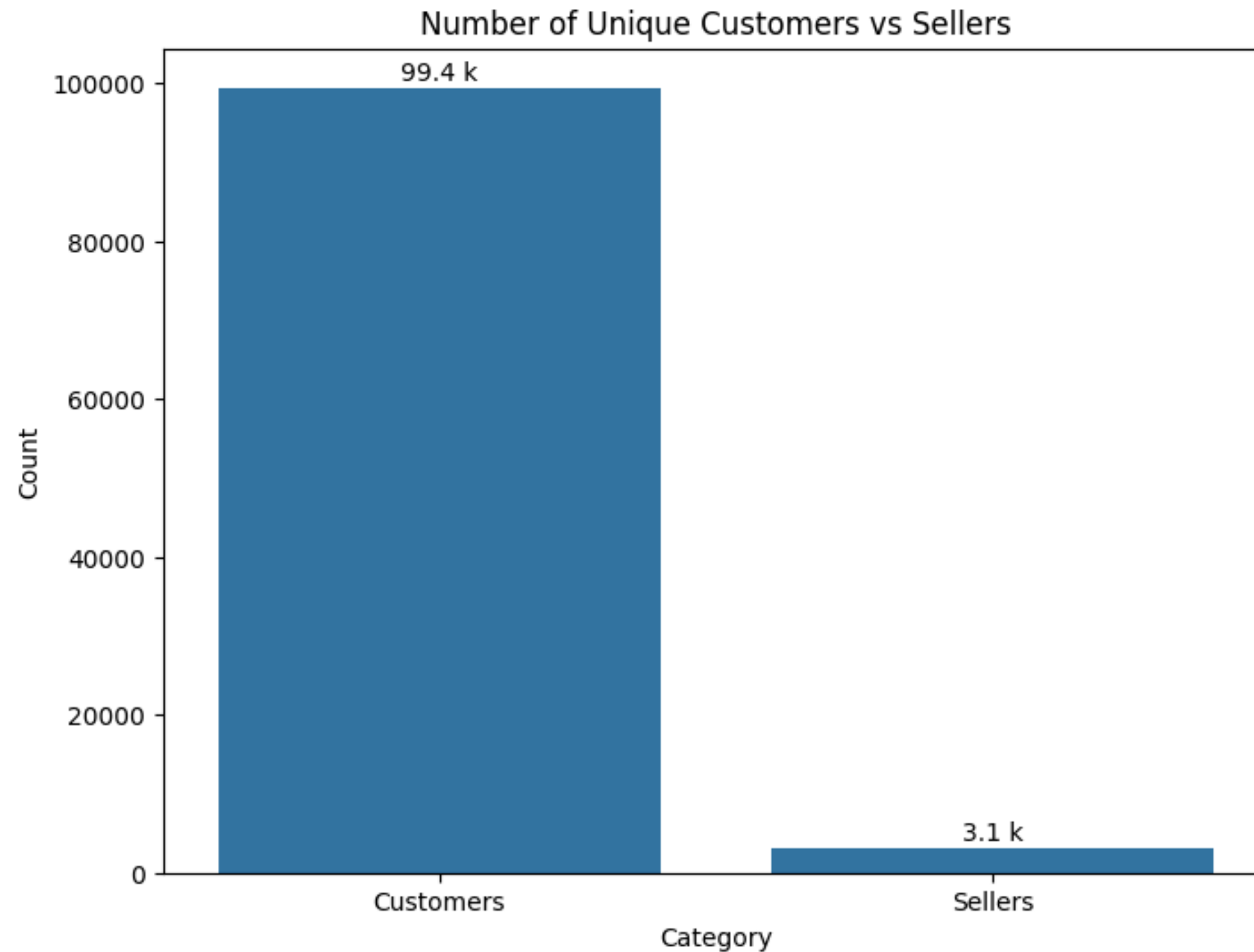


# *Data Overview*

Secara umum terdapat 9 jenis table data, yaitu:

- Customers
  - Geolocation
  - Order\_items
  - Order\_payments
  - Order Reviews
  - Orders
  - Products
  - Product category name translation
  - Sellers
- 

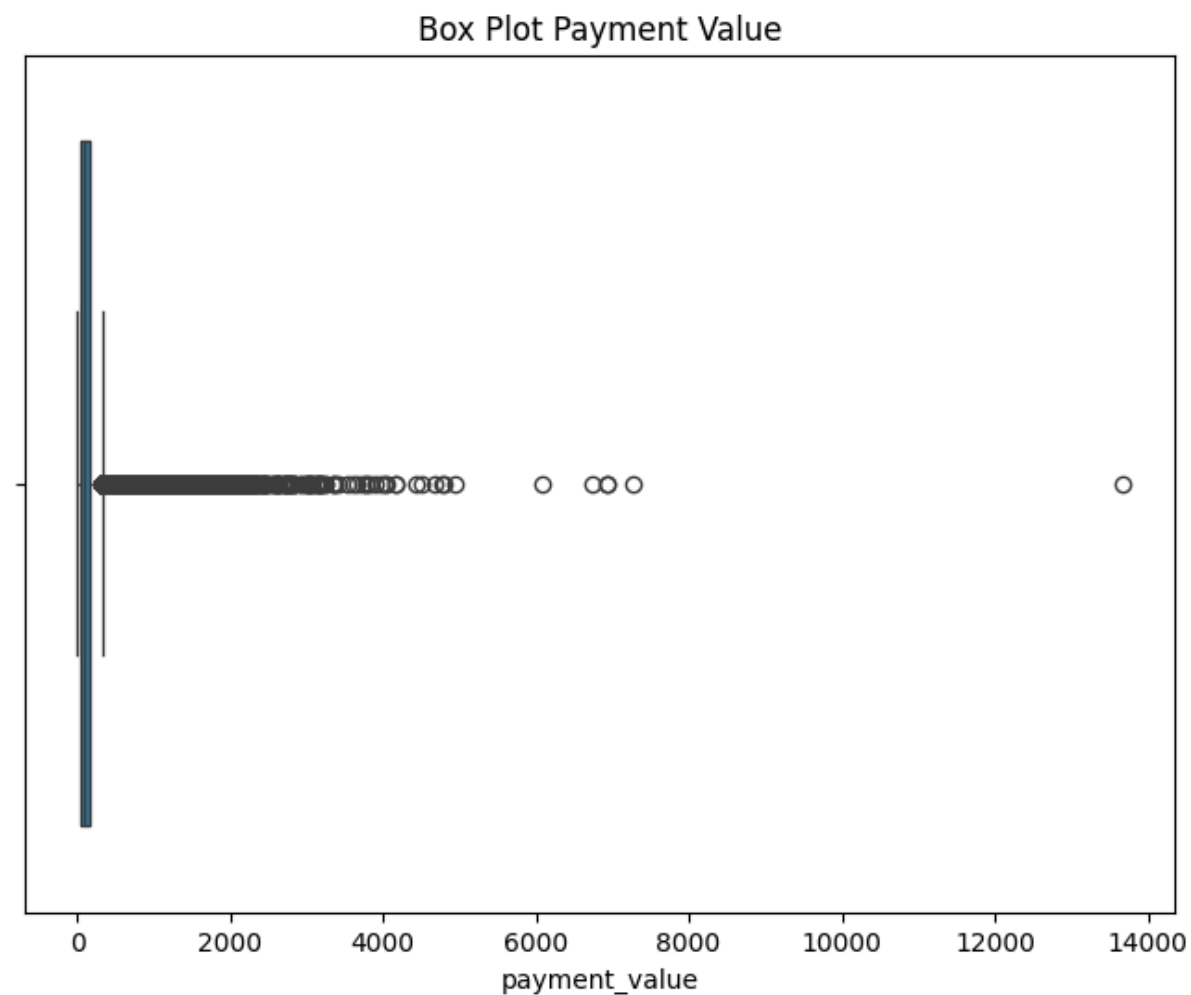
# Ratio Sellers



The seller-to-buyer ratio is **3.02%**, meaning that, on average, one seller **served 32 buyers** throughout the observed period.

# Check Outliers

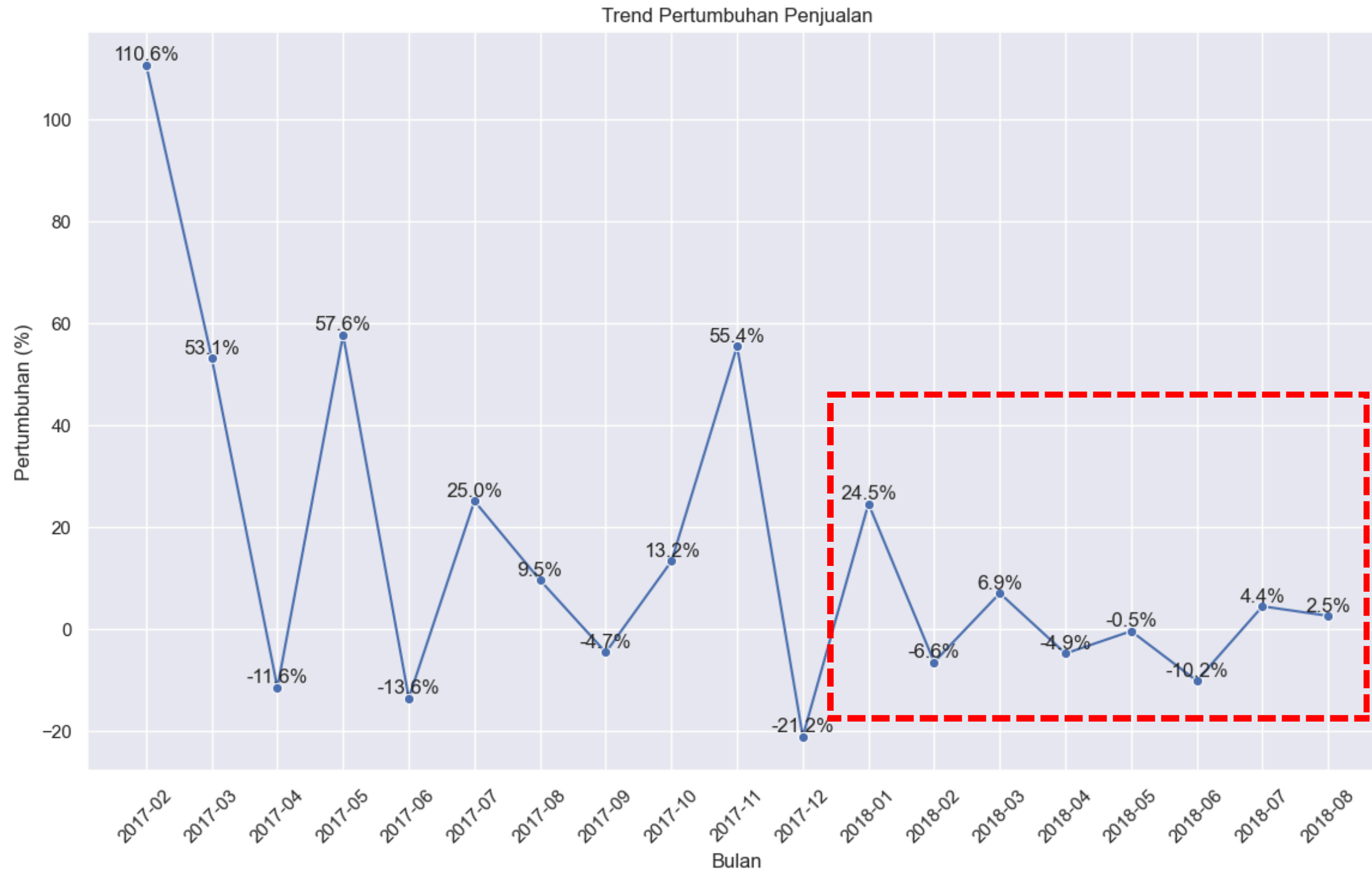
	count	mean	std	min	25%	50%	75%	max
payment_sequential	103886.0	1.092679	0.706584	1.0	1.00	1.0	1.0000	29.00
payment_installments	103886.0	2.853349	2.687051	0.0	1.00	1.0	4.0000	24.00
payment_value	103886.0	154.100380	217.494064	0.0	56.79	100.0	171.8375	13664.08



While the maximum value and Q3 show a large discrepancy due to outliers in payment value, I've chosen not to remove them. This is because, within the context of this dataset, these outliers are not considered abnormal or invalid.

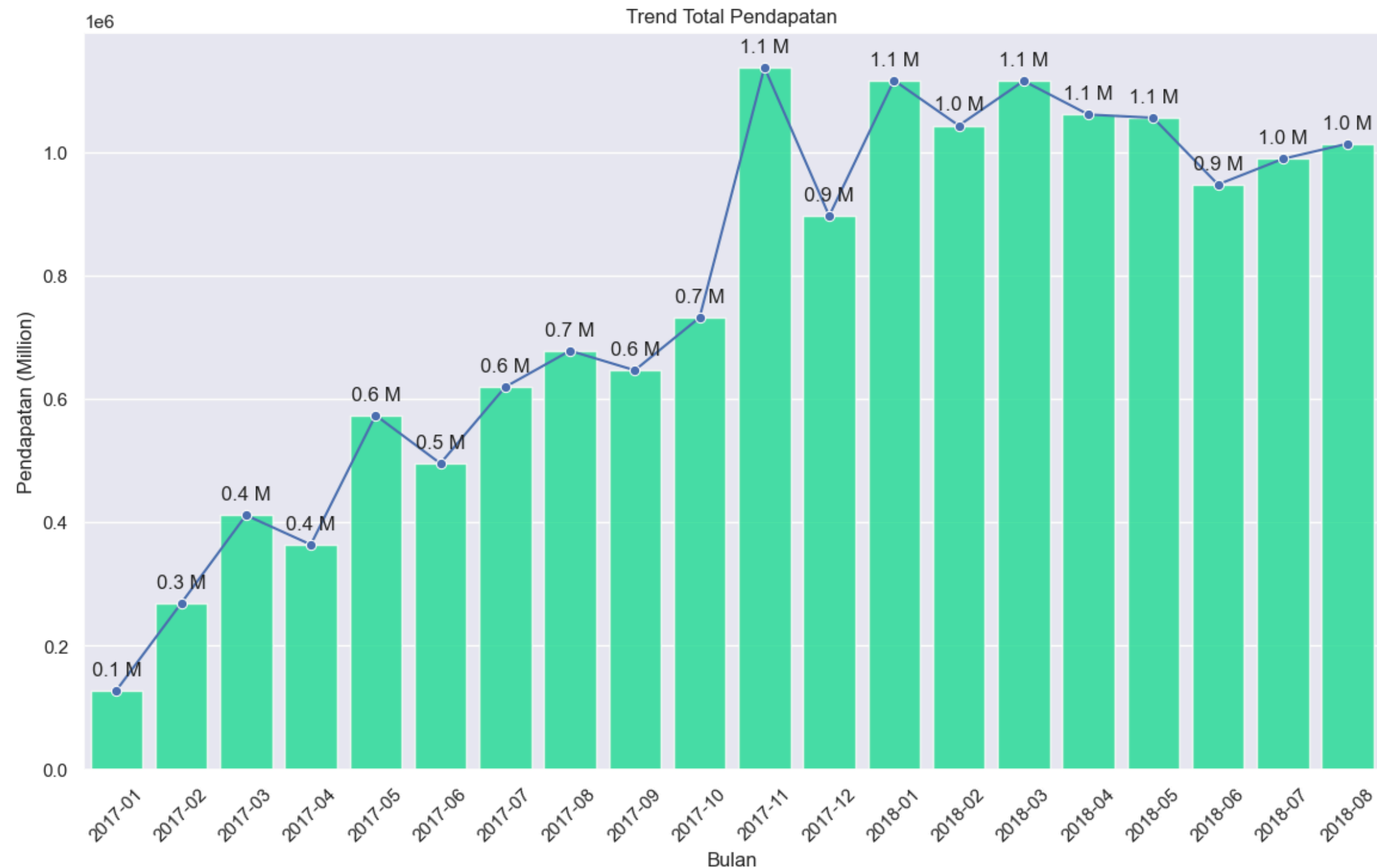


# Sales Growth



Sales growth has been **fluctuating** every month. However, starting from January 2018, it **began to stagnate**.

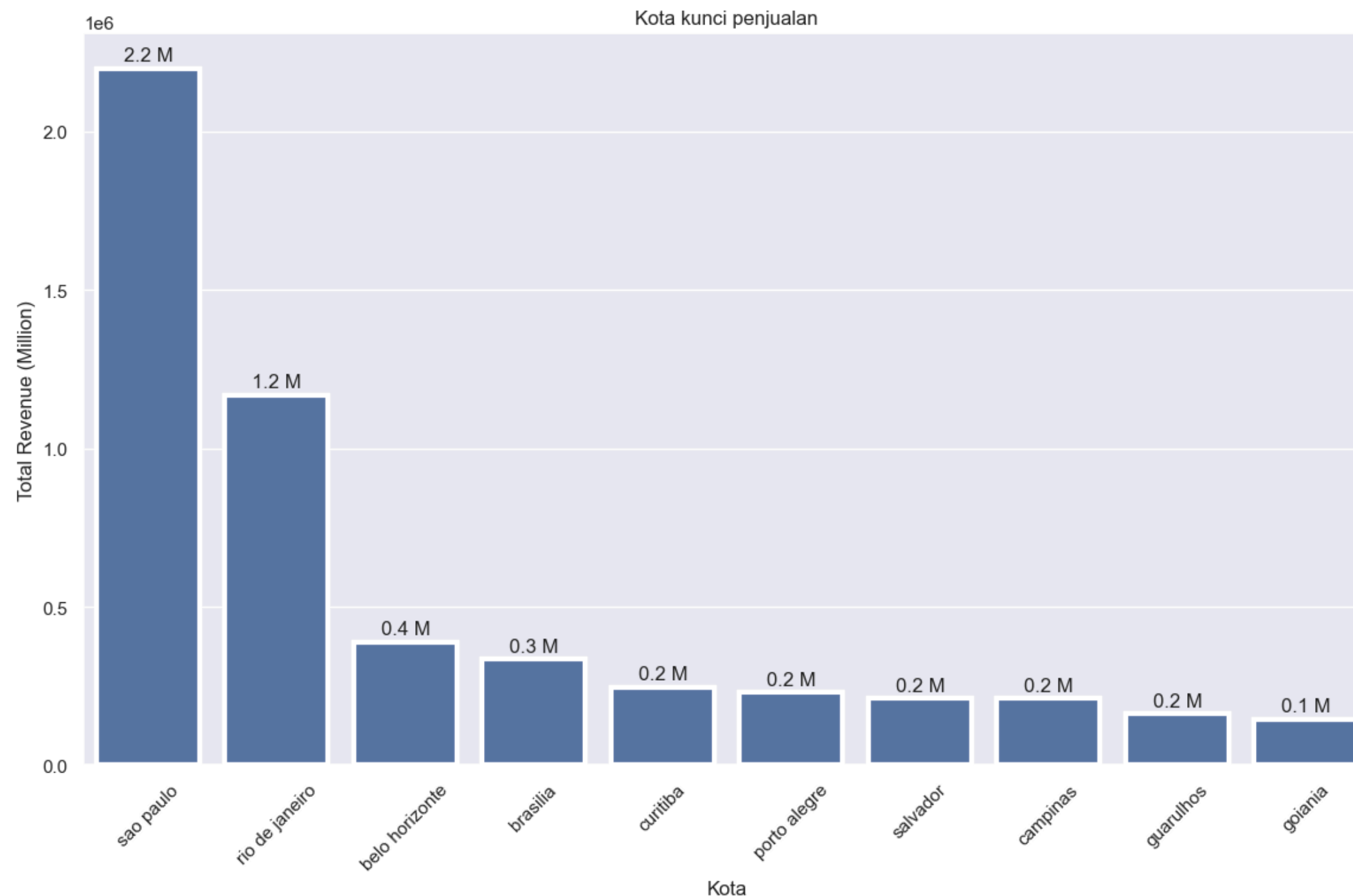
# Revenue



Although there appears to be a decrease in the percentage of revenue growth, it doesn't necessarily mean that the actual revenue has decreased.

The graph shows a **fairly healthy upward trend**, but starting **from January 2018**, there has been a **stagnation in revenue** every month.

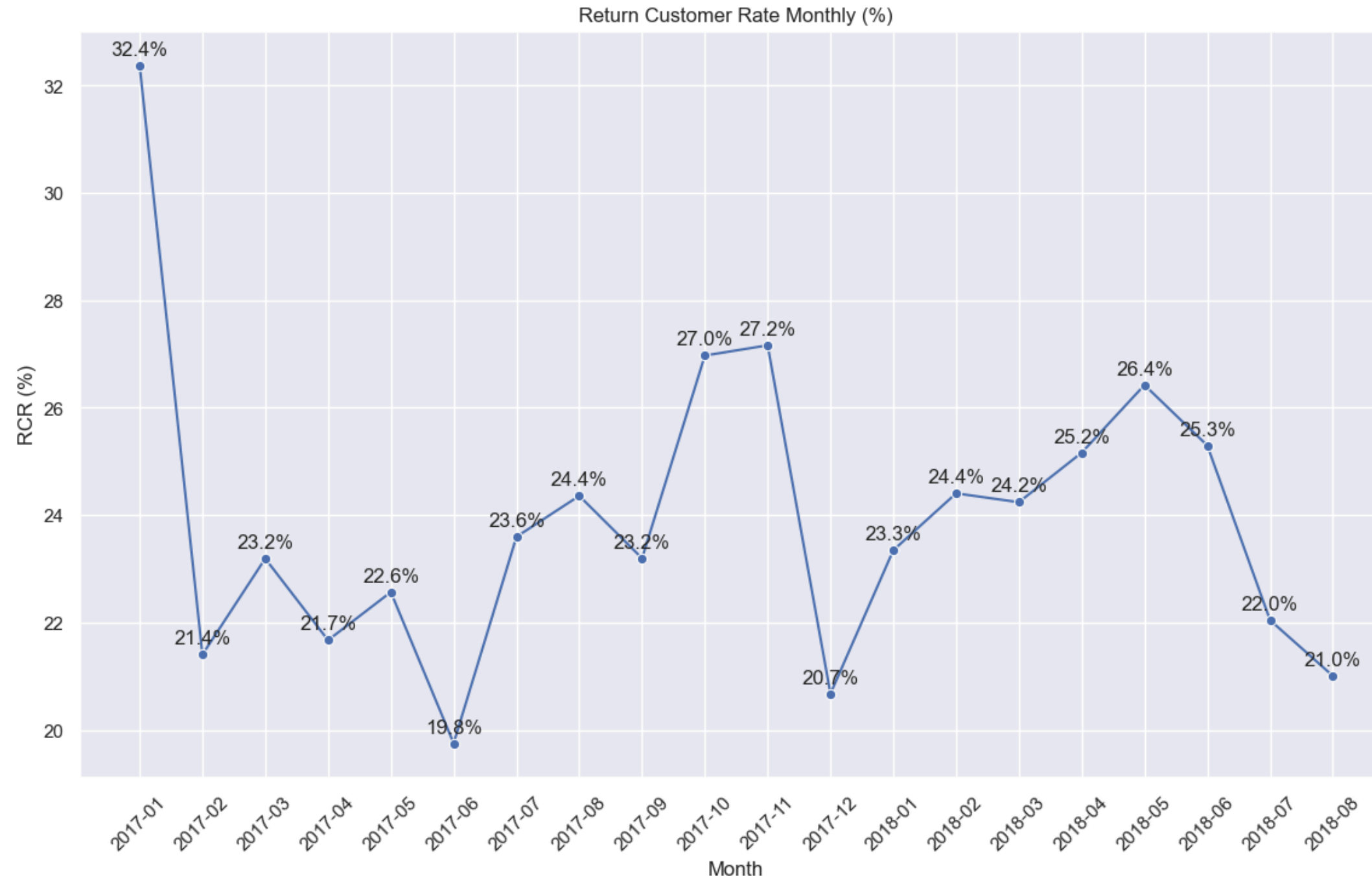
# Top City



**Sao Paulo** has the **highest** and most significant **total sales** compared to other cities. While these **10 cities** could serve as **hubs** for **business** and marketing **strategies**.

If resources are limited, we can further **concentrate** our efforts on just **two cities**: Sao Paulo and Rio de Janeiro.

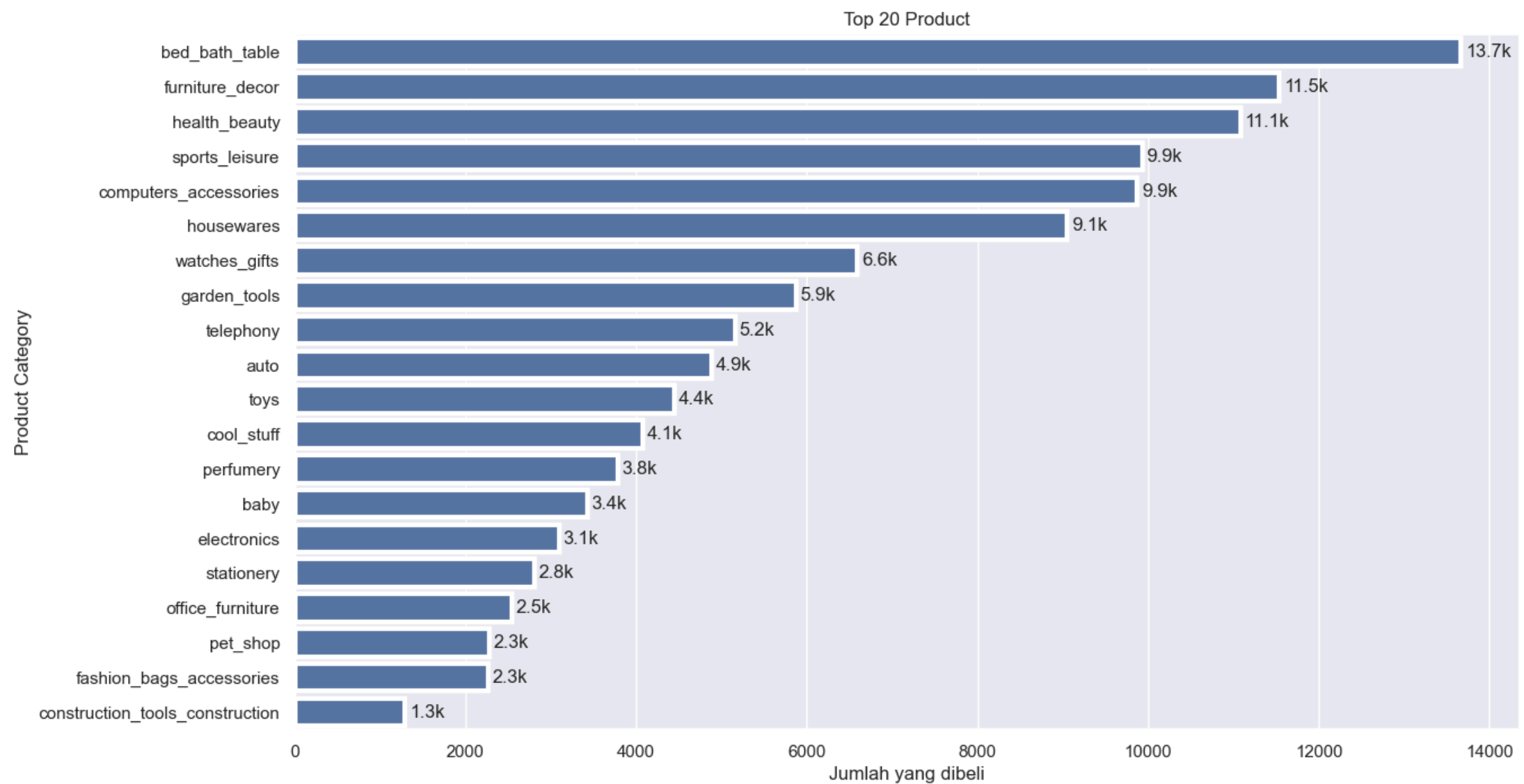
# Return Customer Rate



RCR is a metric that measures the percentage of customers who make subsequent purchases after their initial transaction.

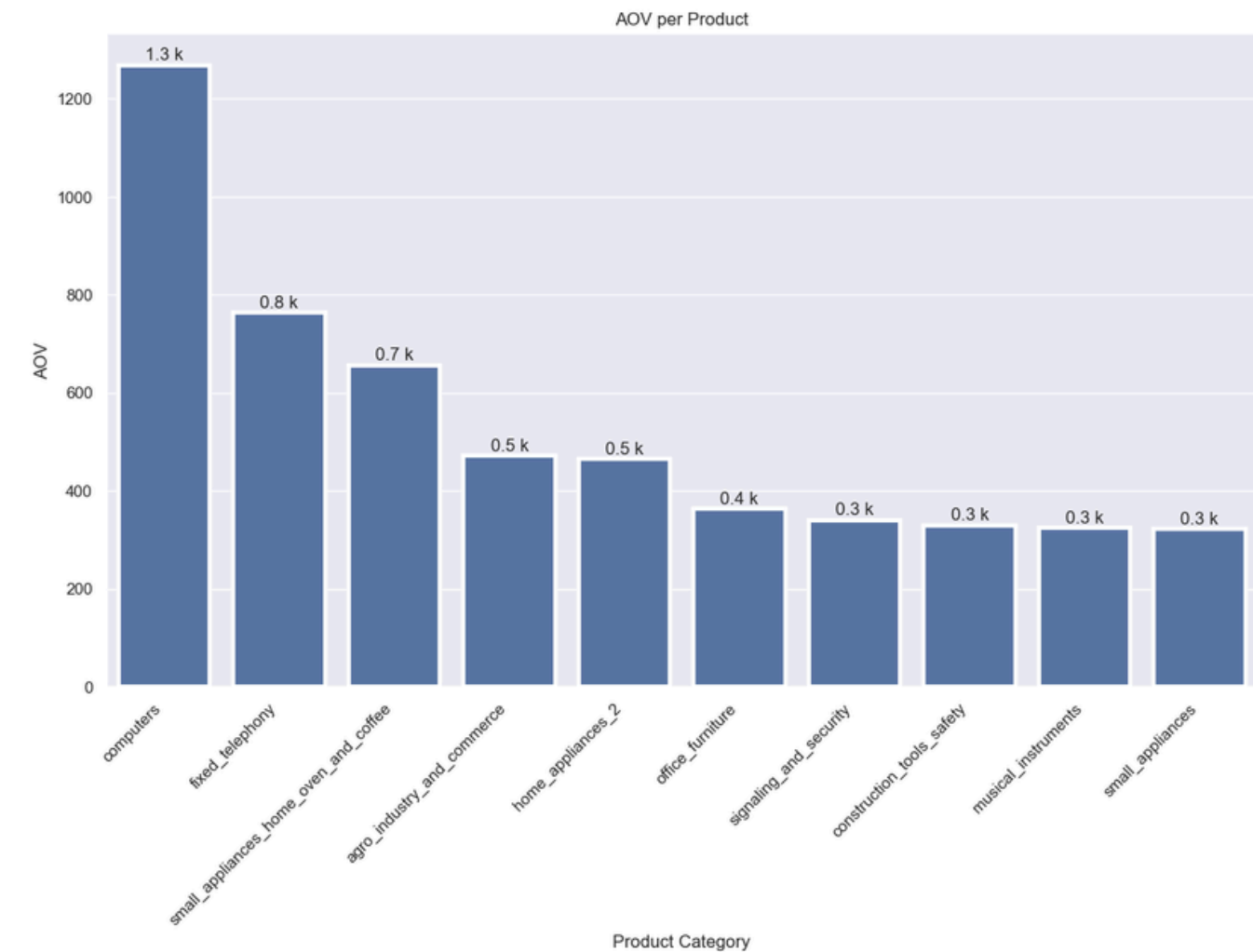
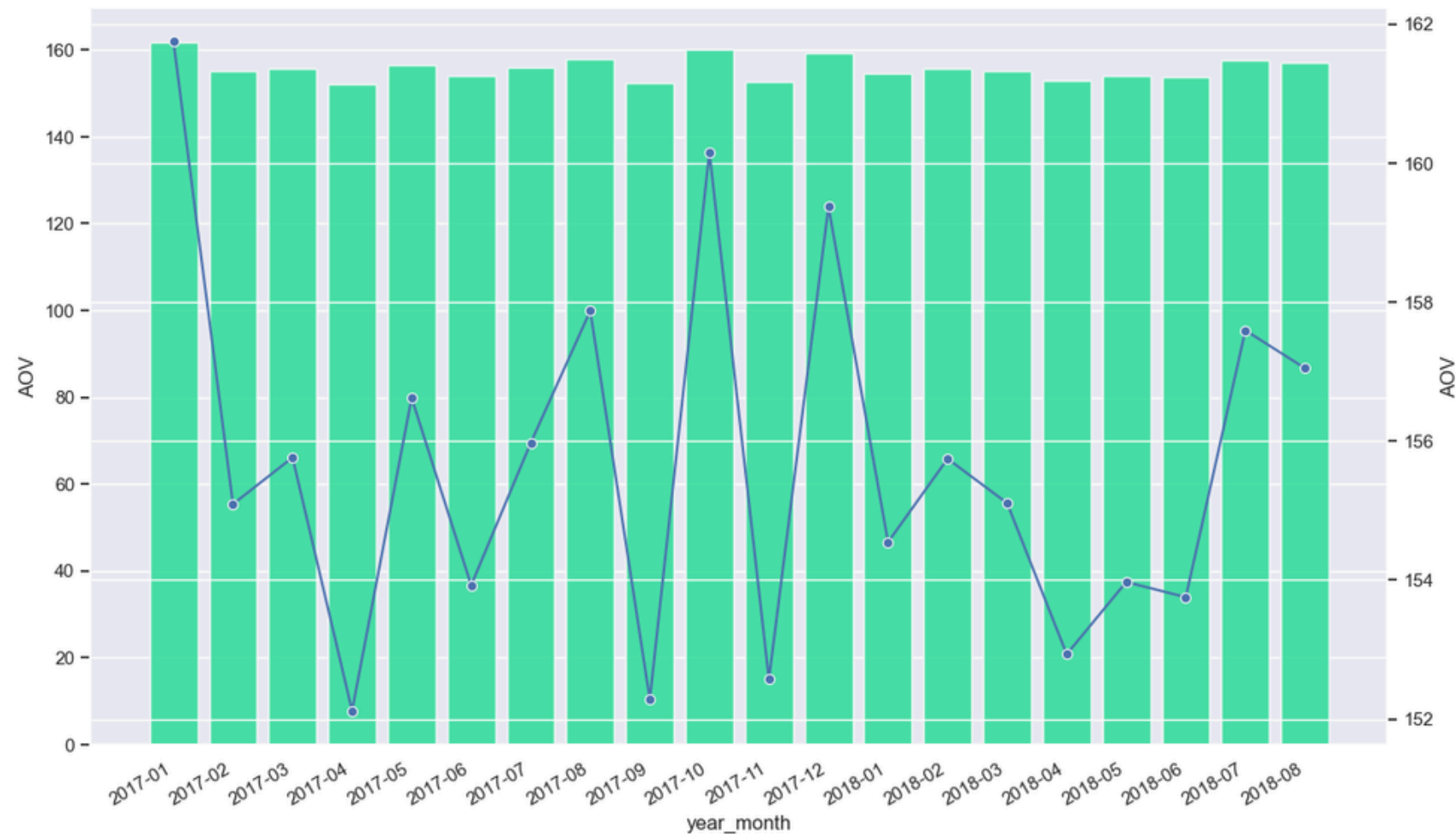
Based on the graph, June and December stand out as the two worst-performing months for our store in terms of RCR.

# Top Product



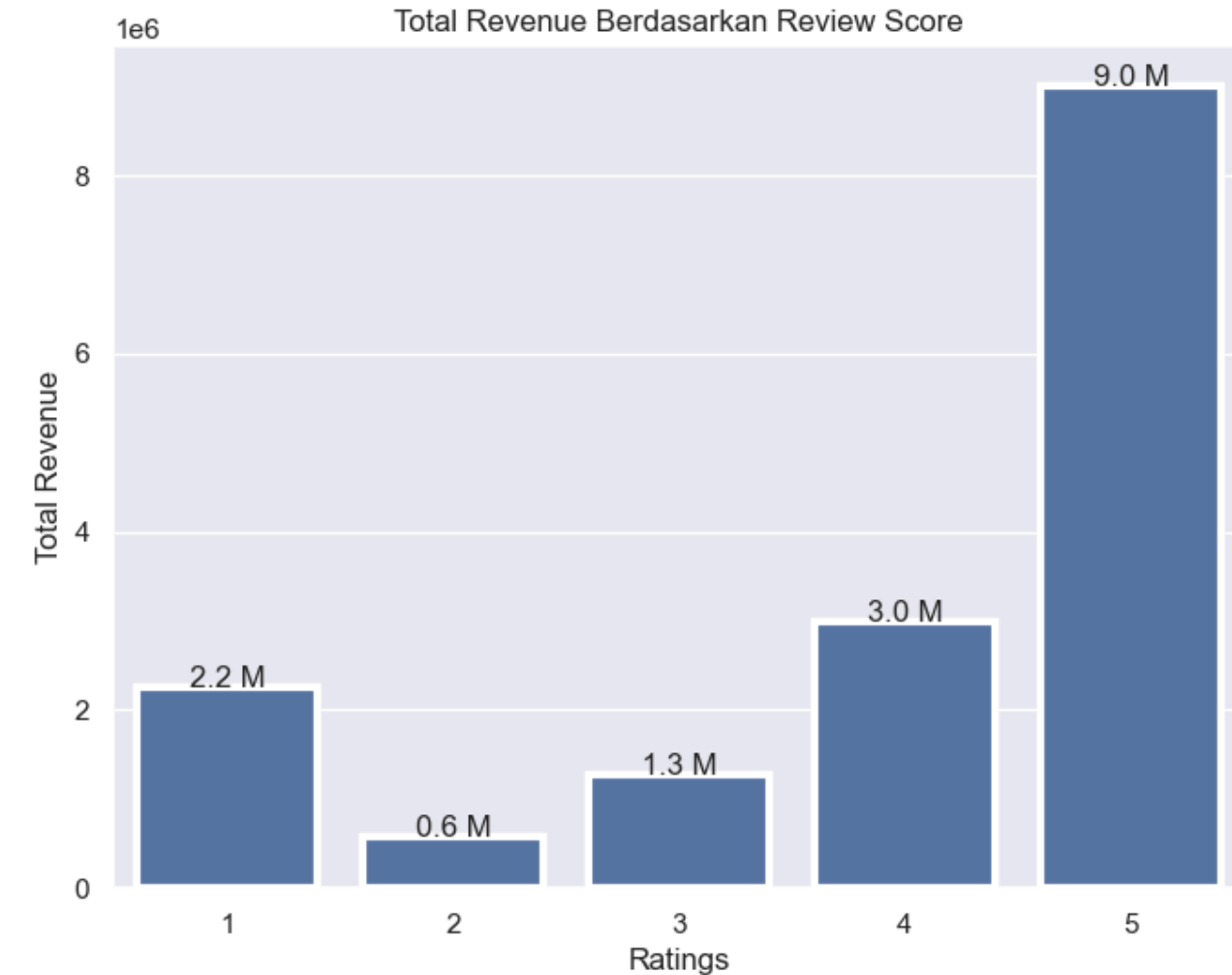
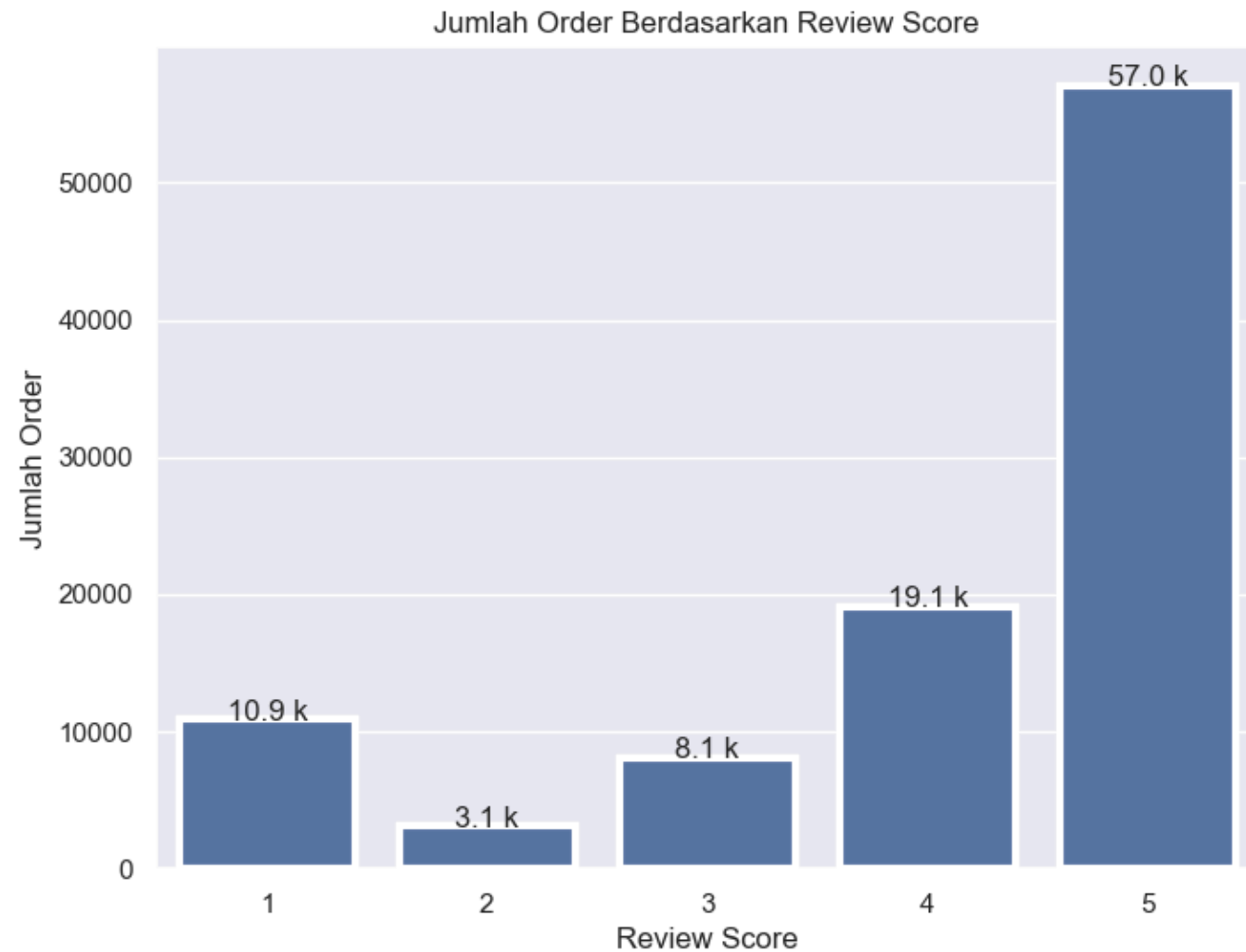
The top 3 product categories (bed\_bath\_table, furniture\_decor, and health\_beauty) account for 35% of the total transaction quantity.

# Avg Order Value



Generally, AOV (Average Order Value) tends to increase at the beginning and end of the year, with "computers" consistently ranking as the product category with the highest AOV.

# Review Score



Based on the graph, it means that highly rated products are in more demand in the market and hence the higher number of orders and revenue generation and customers are fine with spending more for them



# *Section 3*

*Clustering*  
*(RFM Segmentation)*





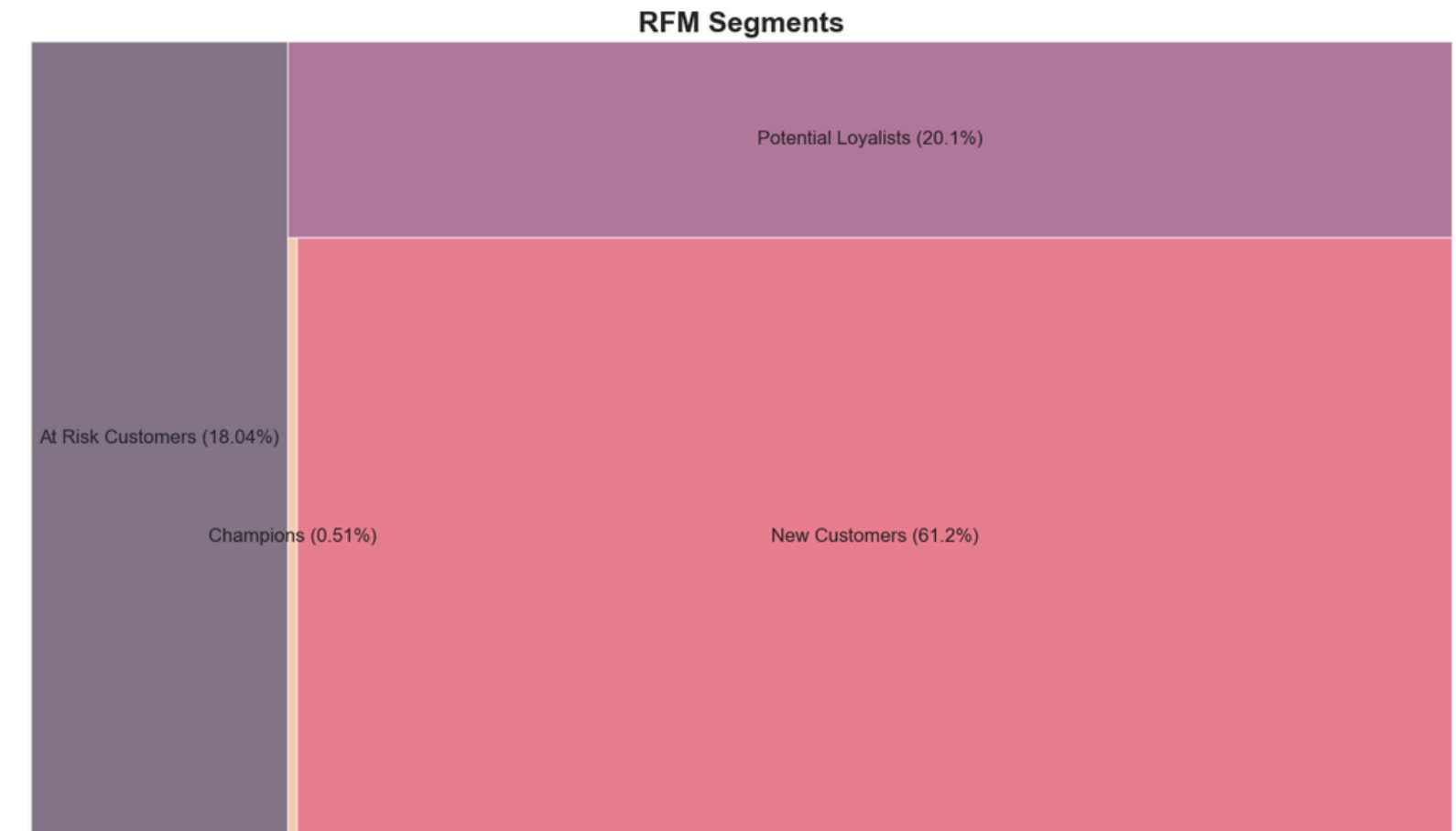
# *Recency, Frequency, and Monetary (RFM)*

- **Recency**  
Refers to when the customer did the most recent transaction using our product
- **Frequency**  
Refers to how often customers do transactions using our product
- **Frequency**  
Refers to how much does a customer spend in our product (add up all the items purchased by the customer and multiply by the price of each item based on additional data)

# RFM Result

	count
rfm_group	
New Customers	60932
Potential Loyalists	20052
At Risk Customers	17947
Champions	510

- **New Customers**  
Recent additions to the customer base with limited purchase history.
- **Potential Loyalists**  
Customers showing promise to become loyal with good Recency and Frequency



- **At Risk Customers**  
Customers who haven't purchased recently (low Recency), potentially with low frequency and/or monetary value as well
- **Potential Loyalists**  
Our most valuable customers with high scores across all RFM metrics



# *Section 4*

## *Business Insight & Recommendation*



# *Business Insight*

Based on the analysis conducted, several key insights have been identified:

- **Sales growth stagnation** -> Sales growth plateaued over the final eight months of 2018.
- **Seasonal RCR decline** -> A consistent pattern of declining Return Customer Rate (RCR) emerges in June and December each year, with an average decrease of 5%.
- **High-value product category** -> Computers commanded the highest Average Order Value (AOV) despite not ranking highest in terms of sales volume or order count.
- **Dominant product categories** -> Bed\_bath\_table emerged as the most popular product category based on order volume, followed by furniture\_decor and health\_beauty, collectively representing 35% of the total transaction quantity.
- **Customer segmentation** -> Four distinct customer segments were identified: at-risk customers, new customers, potential loyalists, and champions.



# *Business Recommendation*

Here are some actionable recommendations based on the insights:

- The company needs to formulate and implement business and marketing strategies promptly to address declining revenue growth, particularly considering the downward trend compared to early 2017. Failure to do so could potentially lead to losses in the coming year. Analyzing cities with the highest sales and leveraging customer segmentation insights are crucial starting points for these strategies.
- Implement targeted interventions, such as promotional campaigns or discounts, during the months exhibiting consistent RCR decline (June and December). This proactive approach aims to bolster RCR throughout the year and mitigate the seasonal dips.



*Thank you!*