

RFM ANALYSIS MODULE 1

AI FOR COMMUNICATION AND MARKETING PROJECT

BY: MARIE E. BASSIL 503962



OBJECTIVE



Understanding customer behavior is crucial for any business to enhance customer satisfaction and loyalty. The use of RFM (Recency, Frequency, Monetary) analysis helps businesses segment their customers based on purchasing behavior.

AI Model Support:

- Predictive modeling using AI can help in forecasting customer behavior and identifying potential high-value customers.
- AI can automate customer segmentation and personalization of marketing efforts, leading to improved customer retention and sales.

Strategic Decision-Making: These insights support broader strategic decisions across the business, ensuring alignment with overall goals.

DATASET

This extensive dataset serves as the foundation for our targeted analysis, enabling the extraction of key insights through the selection of relevant variables tailored to specific business objectives, such as customer behavior analysis through RFM modeling. In the following slide, an explanation is provided of how it was obtained and then reduced to fit our analysis.

- **`ORDER_ID`** --> **UNIQUE IDENTIFIER FOR EACH ORDER.**
- **`CUSTOMER_ID`** --> **UNIQUE IDENTIFIER FOR EACH CUSTOMER.**
- **`ORDER_STATUS`** --> **STATUS OF THE ORDER (E.G., DELIVERED, SHIPPED, PROCESSING).**
- **`ORDER_PURCHASE_TIMESTAMP`** --> **TIMESTAMP WHEN THE ORDER WAS PLACED.**
- **`ORDER_APPROVED_AT`** --> **TIMESTAMP WHEN THE ORDER WAS APPROVED.**
- **`ORDER_DELIVERED_CARRIER_DATE`** --> **DATE WHEN THE ORDER WAS HANDED OVER TO THE CARRIER.**
- **`ORDER_DELIVERED_CUSTOMER_DATE`** --> **DATE WHEN THE ORDER WAS DELIVERED TO THE CUSTOMER.**
- **`ORDER_ESTIMATED_DELIVERY_DATE`** --> **ESTIMATED DELIVERY DATE FOR THE ORDER.**
- **`CUSTOMER_UNIQUE_ID`** --> **UNIQUE IDENTIFIER FOR EACH CUSTOMER (ALTERNATIVE TO `CUSTOMER_ID`).**
- **`CUSTOMER_ZIP_CODE_PREFIX`** --> **ZIP CODE PREFIX OF THE CUSTOMER'S LOCATION.**
- **`CUSTOMER_CITY`** --> **CITY OF THE CUSTOMER'S LOCATION.**
- **`CUSTOMER_STATE`** --> **STATE OF THE CUSTOMER'S LOCATION.**
- **`PAYMENT_SEQUENTIAL`** --> **SEQUENTIAL NUMBER OF THE PAYMENT WITHIN AN ORDER.**
- **`PAYMENT_TYPE`** --> **TYPE OF PAYMENT (E.G., CREDIT_CARD, VOUCHER).**
- **`PAYMENT_INSTALLMENTS`** --> **NUMBER OF INSTALLMENTS FOR THE PAYMENT.**
- **`PAYMENT_VALUE`** --> **VALUE OF THE PAYMENT.**
- **`ORDER_ITEM_ID`** --> **UNIQUE IDENTIFIER FOR EACH ITEM WITHIN AN ORDER.**
- **`PRODUCT_ID`** --> **UNIQUE IDENTIFIER FOR EACH PRODUCT.**
- **`SELLER_ID`** --> **UNIQUE IDENTIFIER FOR EACH SELLER.**
- **`SHIPPING_LIMIT_DATE`** --> **LATEST DATE BY WHICH THE ORDER SHOULD BE SHIPPED.**
- **`PRICE`** --> **PRICE OF THE PRODUCT.**
- **`FREIGHT_VALUE`** --> **FREIGHT VALUE FOR THE SHIPPING OF THE PRODUCT.**
- **`PRODUCT_CATEGORY_NAME`** --> **CATEGORY NAME OF THE PRODUCT.**
- **`PRODUCT_NAME_LENGTH`** --> **LENGTH OF THE PRODUCT NAME.**
- **`PRODUCT_DESCRIPTION_LENGTH`** --> **LENGTH OF THE PRODUCT DESCRIPTION.**
- **`PRODUCT_PHOTOS_QTY`** --> **NUMBER OF PHOTOS AVAILABLE FOR THE PRODUCT.**
- **`PRODUCT_WEIGHT_G`** --> **WEIGHT OF THE PRODUCT IN GRAMS.**
- **`PRODUCT_LENGTH_CM`** --> **LENGTH OF THE PRODUCT IN CENTIMETERS.**
- **`PRODUCT_HEIGHT_CM`** --> **HEIGHT OF THE PRODUCT IN CENTIMETERS.**
- **`PRODUCT_WIDTH_CM`** --> **WIDTH OF THE PRODUCT IN CENTIMETERS.**
- **`PRODUCT_CATEGORY_NAME_ENGLISH`** --> **ENGLISH NAME OF THE PRODUCT CATEGORY.**

DATASET

DESCRIPTION

The final dataset used for analysis is a comprehensive amalgamation of several distinct datasets. The dataset contains customer transaction and comes from "Olist", a Brazilian ecommerce store which has information of 100k orders from 2016 to 2018 made at multiple marketplaces in Brazil where its features allow viewing an order from multiple dimensions: from order status, price, payment and freight performance to customer location, product attributes and finally reviews written by customers.

The merging process was conducted using key identifiers like `customer_id`, `order_id`, and `product_id`, ensuring that relevant data from each dataset was accurately combined.

Main considered variables

- `order_id`: Unique identifier for each order, helping to track individual transactions.
- `customer_unique_id`: Unique identifier for each customer, allowing for customer-level analysis.
- `order_status`: Status of the order (e.g., delivered, shipped, processing), used to filter and analyze completed transactions.
- `order_purchase_timestamp`: Timestamp when the order was placed, critical for calculating recency.
- `product_category_name_english`: The English name of the product category, used to understand customer preferences across different product categories.
- `payment_value`: The total value of the payment for each order, essential for calculating the monetary aspect of RFM analysis.

DESCRIPTIVE STATISTICS AND EXPLORATORY DATA ANALYSIS (EDA)

Descriptive Statistics								
price	103473.00	155.17	216.33	0.00	57.60	100.95	173.40	13664.08
	count	mean	std	min	25%	50%	75%	max

Statistical examination

- **Wide Range of Prices:** The dataset exhibits a broad price range, from 0 to over 13,000 units, indicating a diverse product portfolio.
- **Skewed Distribution:** The distribution of prices is right-skewed, with a few high-priced items significantly affecting the mean.
- **Outliers and Zero Values:** Identified outliers (e.g., high-priced items in specific categories like fixed telephony) and zero-price orders were scrutinized, with zero-price orders being removed to maintain the integrity of the analysis.

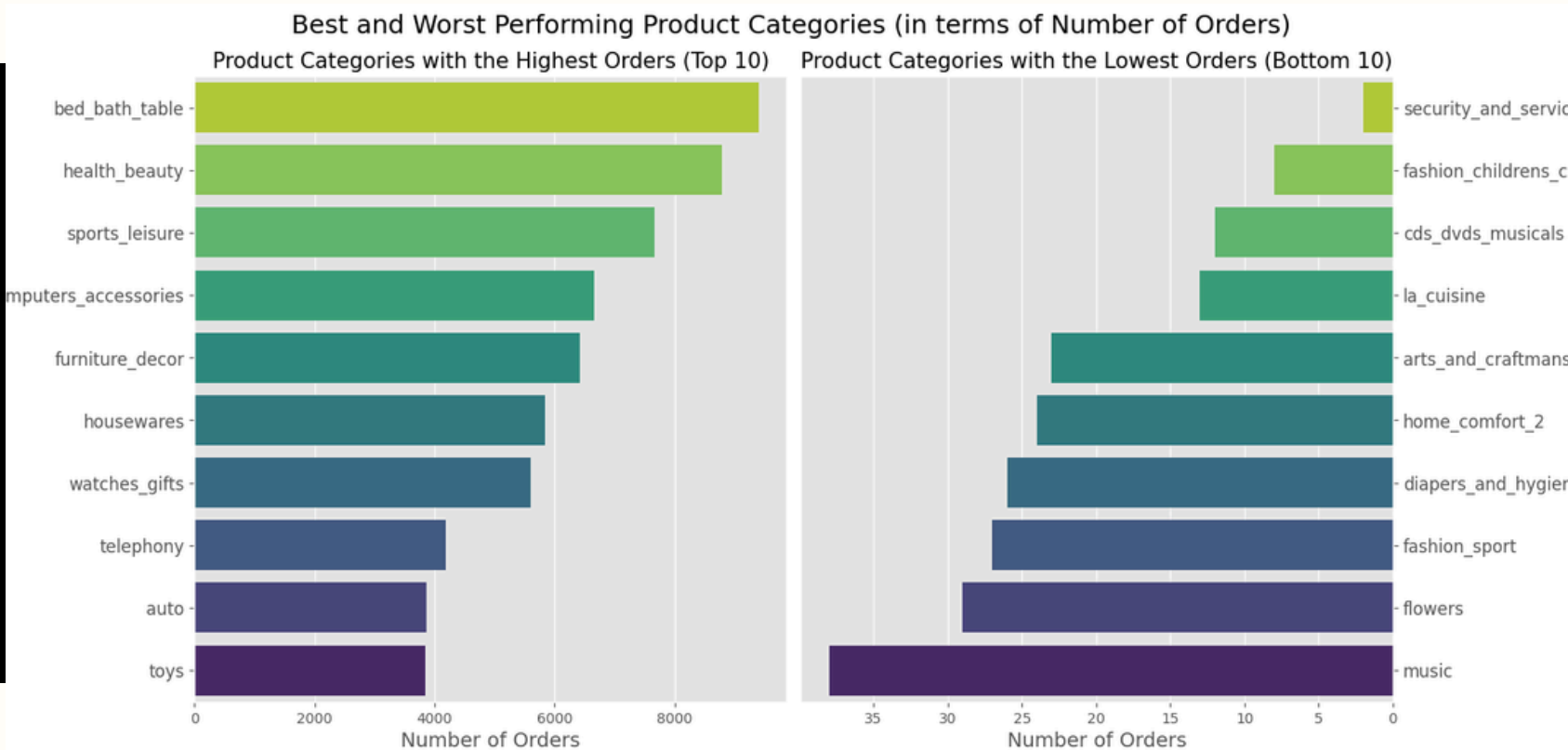
EDA highlights

Order Timings:

- **Peak Times:** Orders peak during the day, particularly around midday to late evening.
- **Low Activity:** Minimal orders are placed in the early morning hours.

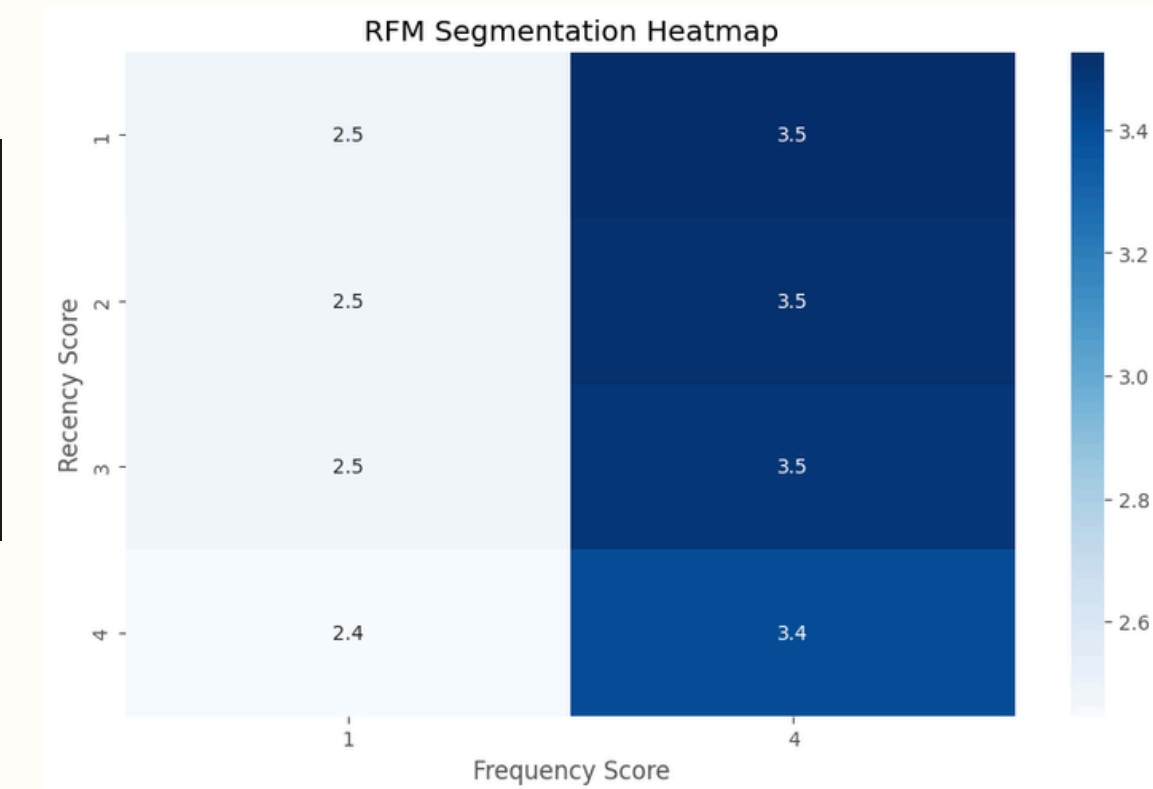
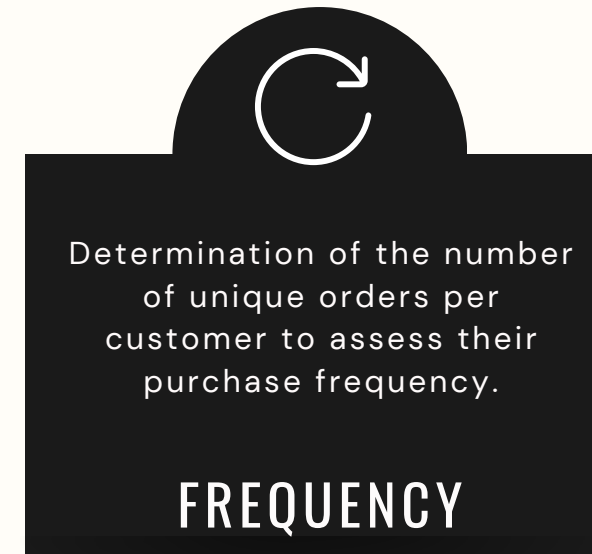
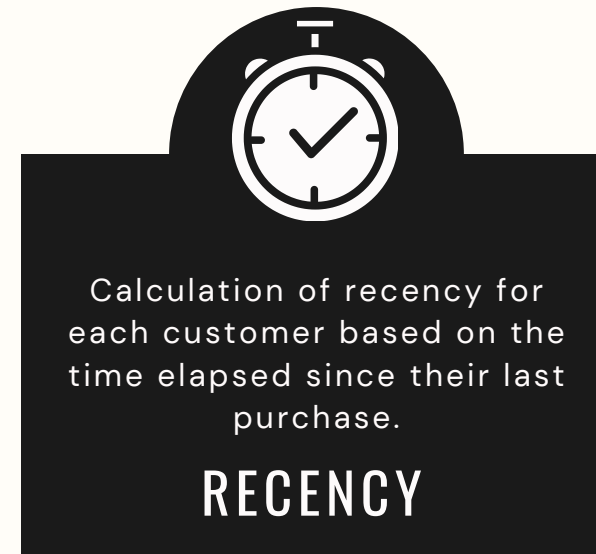
Order Days: Weekday Trends: Monday shows the highest order activity, while weekends experience a dip, indicating varying customer engagement throughout the week.

Product Category Performance: Analysis of order volumes across product categories to identify the top and bottom performers.



RFM ANALYSIS AND CUSTOMER SEGMENTATION

RFM Model Development:



Customers were scored on a scale of 1-4 for each of the RFM metrics based on quartiles. The combined RFM score provides a comprehensive view of customer value.

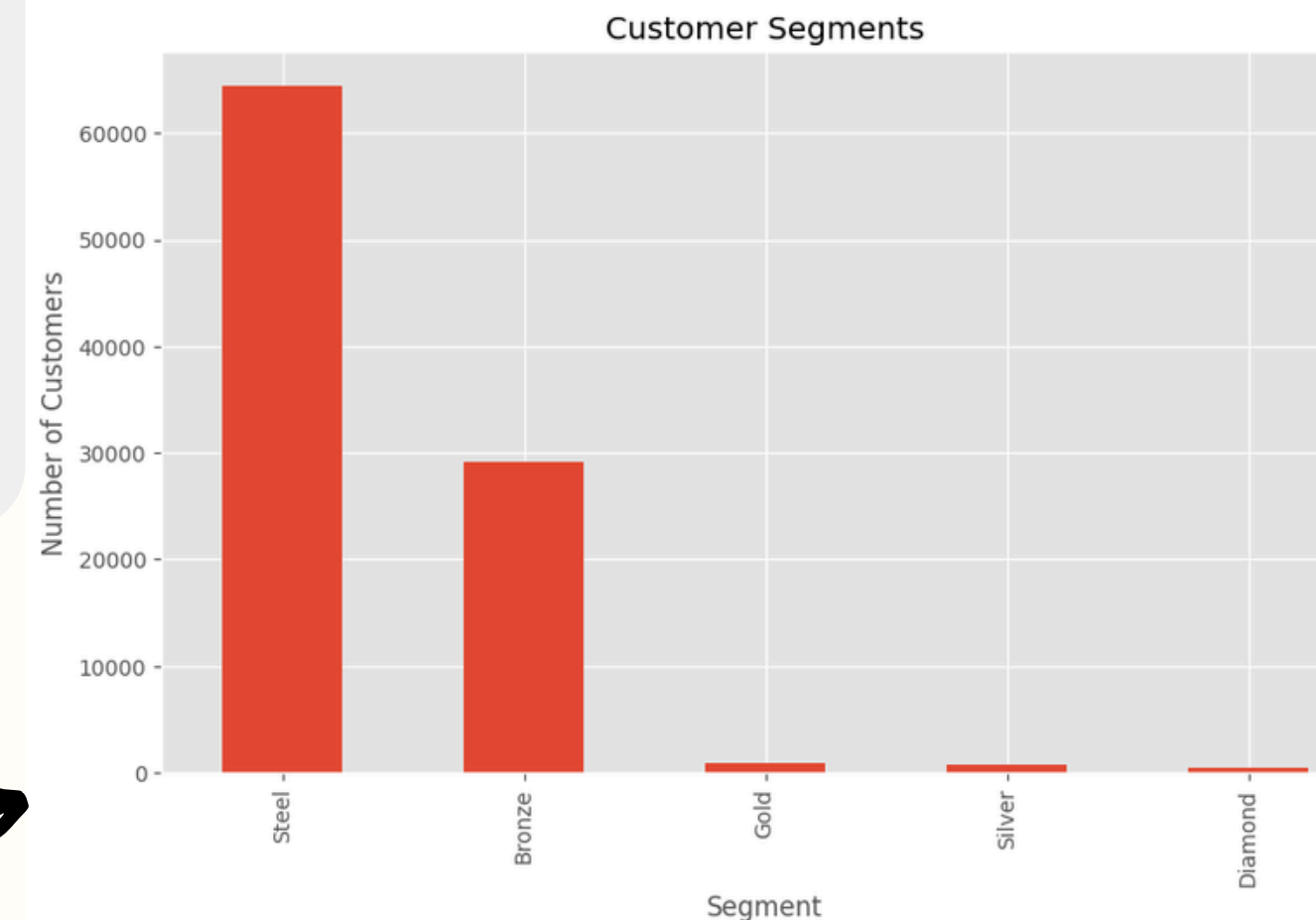
Customers were segmented into groups like Diamond, Gold, Silver, Bronze, and Steel based on their RFM scores.

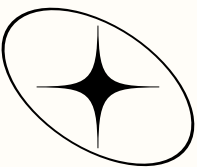
- Diamond: Highest-value customers with recent, frequent purchases and high spending.
- Gold: still valuable and relatively engaged but may not spend as much or purchase as frequently as Diamond customers.
- Silver Segment: moderately engaged, with average spending and purchase frequency.
- Steel: Least recent, infrequent purchases and low spending.

RFM SCORING

CUSTOMER SEGMENTATION

SEGMENT CHARACTERISTICS





CRITICAL COMMENTARY

Product Performance:

- Identified best- and worst-performing product categories, offering clear guidance on where to focus marketing, inventory, and product development efforts.



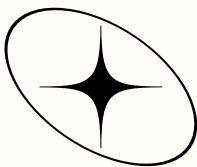
Customer Segmentation Insights:

- The RFM analysis successfully segmented customers into distinct groups, revealing a concentration of customers in lower-value segments (Steel and Bronze) and identifying a small, high-value segment (Diamond) that significantly contributes to revenue.



Engagement Trends:

- The analysis highlighted peak engagement times (e.g., high order volumes on Mondays and during certain hours), which can be leveraged for targeted marketing efforts.



PRACTICAL IMPLICATIONS

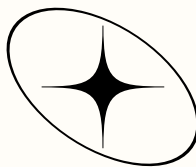
Targeted Marketing Strategies:

- High-Value Segments: Develop personalized and loyalty-driven marketing strategies for the Diamond and Gold segments to retain and further maximize their lifetime value.
- Low-Value Segments: Implement re-engagement campaigns for Bronze and Steel segments, with the aim of moving them up the value chain through targeted promotions and offers.



Operational Adjustments:

- Resource Allocation: Allocate resources more efficiently by focusing on peak ordering times and high-performing product categories to maximize ROI.
- Inventory Management: Use insights from product performance to optimize inventory levels, ensuring that high-demand products are well-stocked, while reconsidering the investment in low-performing items.



RECOMMENDATION

- Regularly update the RFM analysis to track shifts in customer behavior and adjust strategies accordingly.
- Use the insights on customer preferences to identify cross-selling and upselling opportunities, especially among the Gold and Silver segments.



- **The RFM analysis reveals key customer segments, enabling targeted marketing and improved customer retention. Focusing on high-value customers will drive growth, while optimizing resources based on these insights enhances efficiency and profitability. Regular updates will keep strategies aligned with evolving customer needs, ensuring sustained business success.**



AI FOR COMMUNICATION AND MARKETING PROJECT

**THANK
YOU!**

BY: MARIE E. BASSIL 503962

