Predict Customer Personality to Boost Marketing Campaign by Using Machine Learning

Rakamin
Academy

Created by:
Bintang Phylosophie
bintangphy@gmail.com
Bintang Phylosophie

A results-driven Data Scientist and Data Analyst with a strong foundation in machine learning, data analysis, and visualization. Proficient in Python, SQL, and advanced analytics tools such as Power BI, and Looker Studio. Experienced in handling large datasets, optimizing queries, and building data-driven solutions through project-based internships in several company. Holds a Data Science certification from Rakamin Academy with hands-on expertise in statistical modeling, predictive analytics, and interactive dashboard creation. Passionate about data storytelling and leveraging insights to drive business decisions.

Rakamin Academy

Background

A company can develop rapidly when it knows its customers' personality behavior, so that it can provide better services and benefits to customers who have the potential to become loyal customers. By processing historical marketing campaign data to improve performance and target the right customers so they can make transactions on the company's platform, from this data insight our focus is to create a cluster prediction model to make it easier for companies to make decisions..

Dataset & Business Understanding

Dataset Information:
This dataset contains customer information related to demographics, purchasing behavior, and marketing campaign responses. It includes details such as age, education, marital status, income, and household composition. Additionally, it tracks spending across different product categories, purchase channels, and engagement with marketing campaigns. The dataset also records customer complaints and recent interaction. This dataset is from a fictional company in 2012-2014.

Attribute Information:
- **Identifier**
  `ID` is unique ID that each of which is an ID for each customer and a column called `Unnamed: 0` which seems to function more as an index
- **Company Goals:**
  The goal of the company is to enhance customer relationship management by developing data-driven marketing strategies that improve engagement, increase sales, and optimize resource allocation based on customer segmentation insights.
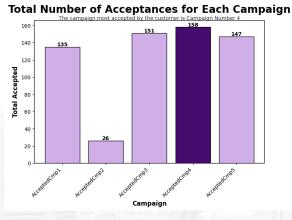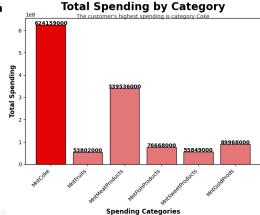
- **Problem:**
  The company wants to better understand its customer base by identifying distinct customer segments based on demographic information, purchasing behavior, and engagement with marketing campaigns. Currently, marketing strategies are applied uniformly across all customers, leading to suboptimal targeting and lower campaign effectiveness. By implementing a clustering model, the company aims to segment customers into meaningful groups to personalize marketing efforts and improve customer retention.
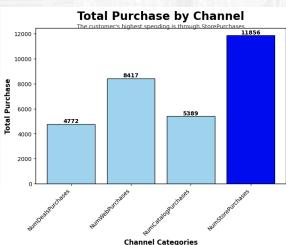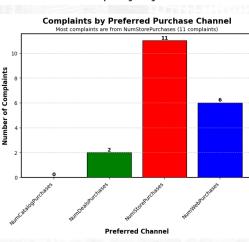- **Objectives:**
  - **Segment Customers** – Use clustering techniques to group customers based on demographics, spending patterns, and engagement with marketing campaigns.
  - **Understand Customer Behavior** – Identify key characteristics of each segment, such as spending habits, preferred purchasing channels, and responsiveness to marketing efforts.
  - **Optimize Marketing Strategies** – Develop personalized marketing campaigns tailored to different customer segments to improve conversion rates and customer satisfaction.
  - **Enhance Customer Retention** – Identify high-value customers and at-risk customers to implement targeted retention strategies.
  - **Improve Business Efficiency** – Allocate marketing resources effectively by focusing on the most profitable and responsive customer segments.

Insight and Analysis

Total Number of Acceptances for Each Campaign

The campaign most accepted by the customer is Campaign Number 4

Total Spending by Category

The customer's highest spending is category Coke

Total Purchase by Channel

The customer's highest spending is through StorePurchases

Complaints by Preferred Purchase Channel

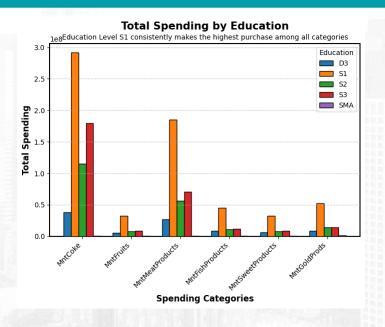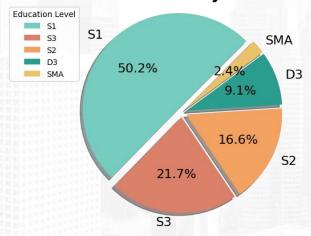Most complaints are from NumStorePurchases (11 complaints)

- **Campaign 4** had the highest number of acceptances (158), followed by Campaign 3 (151), Campaign 5 (147), and Campaign 1 (135). Campaign 2 had the lowest number of acceptances (26), which may indicate an issue with the campaign's effectiveness.
- The highest spending category was **"MntCoke"** (Rp 624,159,000), followed by "MntMeatProducts" (339,536,000). Other categories, such as "MntFruits" (Rp 53,802,000) and "MntSweetProducts" (Rp 55,849,000), had significantly lower spending. This suggests that the customer prioritizes beverages and meat products over other food categories.
- The majority of purchases were made through **store purchases** (11,856), followed by web purchases (8,417). This indicates that customers prefer in-store shopping over online or catalog-based purchases.
- Store purchases had the highest number of complaints (11), followed by web purchases (6).The high number of complaints in store purchases **aligns** with the fact that it is the most used channel, This suggests that higher engagement leads to more reported issues.

Total Spending by Education

Education Level S1 consistently makes the highest purchase among all categories

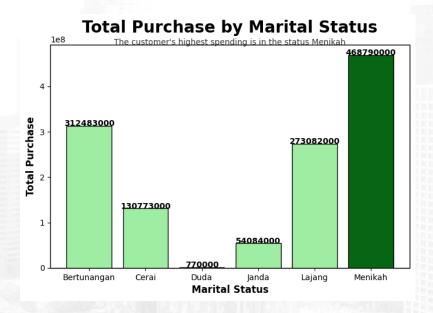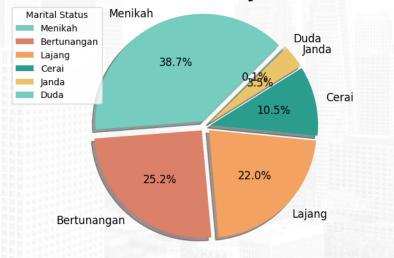**Customer Distribution by Education Level**

- Customers with an **S1 (Bachelor's)** degree spend the most in all categories, S3 (Doctorate) and S2 (Master's) degree holders also spend significantly, but less than S1. Customers with SMA (High School) education spend the least across all product categories. This is align with 50.2% of total customer from S1, followed by 21.7% from S3 holders.
- Higher education correlates with higher spending. Businesses should focus on educated customers (especially S1 holders) for premium products. Develop budget-friendly product lines for lower education groups (SMA) who have lower spending power.
- **Coke** accounts for the largest share of spending, followed by Meat Products. Fish, Fruits, and Sweet Products have significantly lower spending. Gold Products also show substantial spending, suggesting interest in premium/luxury items. Invest in promotions for beverages (Coke) and protein-rich foods (Meat & Fish), as they are the most purchased items across all education levels.
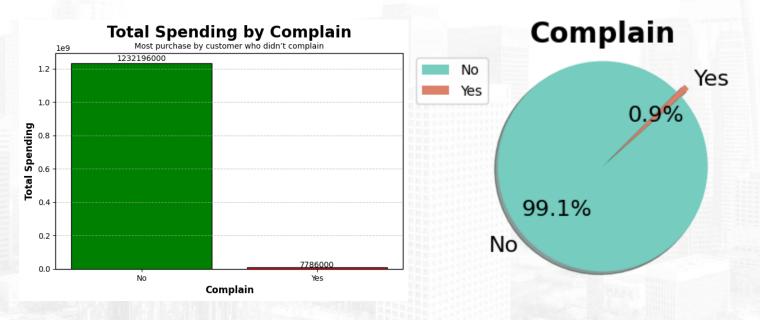
Insight and Analysis

Total Purchase by Marital Status

The customer's highest spending is in the status Menikah

Customer Distribution by Marital Status

- **Menikah (married)** individuals had the highest spending (Rp 468,790,000), followed by Bertunangan (engaged) individuals (Rp 312,483,000). Married individuals contribute the highest spending, aligning with their highest representation in the customer base. Lanjangn (single) (Rp 273,082,000) and Cerai (divorced) (Rp 130,773,000) individuals had lower spending.
- Campaigns targeting married and engaged individuals may yield higher revenue.
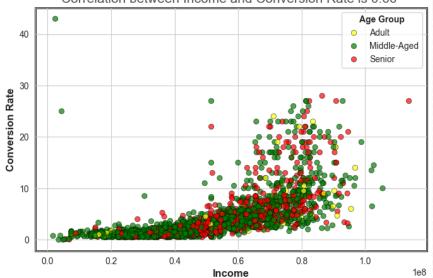
Total Spending by Complain — Most purchase by customer who didn't complain. Bar chart showing No: 1232196000, Yes: 7786000.

Complain — Pie chart: No 99.1%, Yes 0.9%.

- Customers **who did not complain** spent significantly more (Rp 1,232,196,000) compared to customers who did complain (Rp 7,786,000).
- **99.1%** of customers did not complain, while only 0.9% complained. This indicates that the majority of customers are satisfied, or at least do not formally register complaints.
- The data could also imply that customers who face issues tend to stop purchasing instead of lodging complaints, meaning the actual dissatisfaction rate could be higher than 0.9%.
- Since only 0.9% of customers complain, businesses should encourage more feedback to identify hidden dissatisfaction.
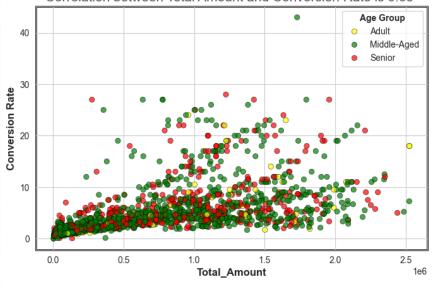
Conversion Rate vs. Income (Colored by Age Group)
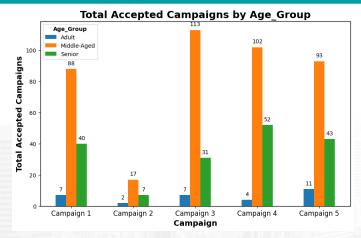Correlation between Income and Conversion Rate is 0.66

Conversion Rate vs. Total Amount (Colored by Age Group)
Correlation between Total Amount and Conversion Rate is 0.63
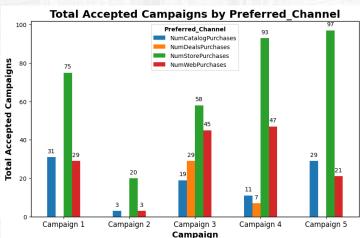
- The correlation coefficients of Income (0.66) and Total Amount (0.63) with Conversion Rate suggest **a moderate to strong positive** correlation between both Income and Total Amount (total spending) and Conversion Rate.
- The distribution appears evenly spread across different age groups, with **no single group dominating the trend**.
- The relationship is not perfectly linear; higher incomes and total amount show more variability in conversion rates.

Insight and Analysis

Total Accepted Campaigns by Age_Group

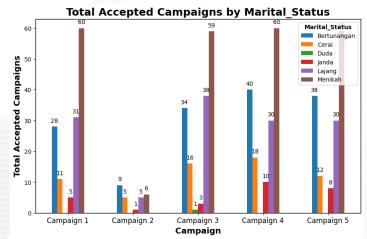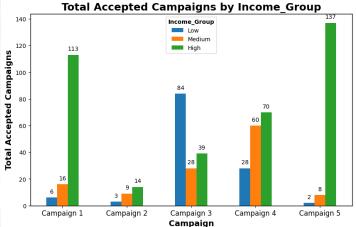Total Accepted Campaigns by Preferred_Channel

- The middle-aged group is the most responsive to campaigns, making them the primary target audience. Seniors also engage significantly, meaning campaigns should still consider them. Adults show very low engagement, suggesting they are less interested or the campaigns are not aligned with their preferences.
- Campaign 3 and Campaign 4 have the highest participation, meaning they were either well-designed, relevant, or well-promoted. Campaign 2 had the lowest participation, indicating potential issues in marketing, reach, or relevance.
- Focus on tailoring campaigns towards the middle-aged and senior age groups. Investigate why adults have low participation and adjust campaign messaging accordingly. Analyze what made Campaign 3 and Campaign 4 successful and replicate those strategies.

- Physical store purchases are the preferred channel across all campaigns. Marketers should focus on enhancing in-store experiences. Continue strong in-store promotions, as this is the dominant purchasing behavior.
- Online purchases and promotional deals work well together, especially in Campaign 3 and 4. Digital marketing strategies and promotional campaigns can improve engagement. Enhance digital and promotional campaigns to engage web shoppers and deal-seekers..
- Catalog purchases remain relevant in some cases (Campaigns 1, 3, and 5) but show inconsistency. This channel may need optimization.
- Campaign 2 was the least successful across all channels, suggesting potential weaknesses in its marketing approach.

Total Accepted Campaigns by Marital_Status

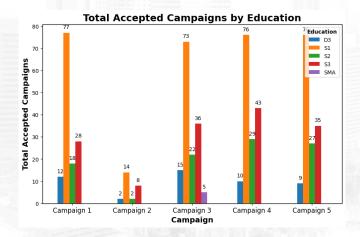Total Accepted Campaigns by Income_Group

- Married individuals are the most responsive to campaigns, campaigns should be optimized for this demographic. Personalized messaging and incentives tailored to family needs could improve conversions further.
- Engaged and single individuals also show strong engagement, making them potential key demographics for future campaigns.
- Divorced, widowed, and separated individuals participate less, suggesting campaigns may need adjustments to appeal to them. Customizing offers to better meet their needs (e.g., financial security, lifestyle products) could enhance engagement.
- Combining marital status data with other demographic insights like age, income, and interests can refine targeting.

- The high-income group has the highest number of accepted campaigns across all five campaigns. suggesting a strong correlation between financial capacity and campaign engagement. campaigns targeting them should emphasize luxury, exclusivity, and premium services.
- The Medium-income group is somewhat responsive but shows varied engagement across campaigns. but certain campaigns (like Campaign 3 and 4) performed well.. Testing different pricing structures, installment plans, or value-based promotions can increase engagement.
- The Low-income group has very limited engagement, indicating potential barriers like affordability, lack of interest, or mismatched targeting. assess whether pricing, messaging, or product offerings are misaligned with their needs.
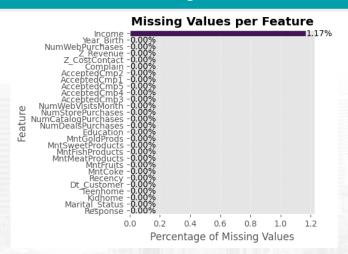
Total Accepted Campaigns by Education

- The S1 and S3 groups dominate the accepted campaigns, meaning those with a Bachelor's or Doctorate degree are more engaged in campaigns. This suggests that higher education levels correlate with higher participation rates.
- The low participation rates in SMA (high school graduates) and D3 (diploma holders) suggest that campaigns may not be targeting or appealing to these groups effectively.
- Campaign 1 and Campaign 4 had the highest participation rates, meaning they were well-structured or relevant to the audience. Campaign 2 had the lowest acceptance, indicating that it was either poorly marketed or less relevant to participants.
- Focus campaigns on individuals with a Bachelor's or higher education, as they show the most engagement. Reassess the approach for SMA and D3 groups to find ways to improve their participation.
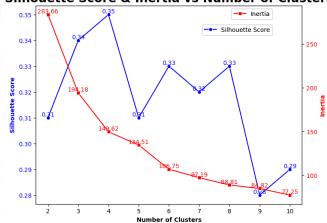
Data Preprocessing

Missing Value

Missing Values per Feature

- `Income` is only one features that contain missing value, and after long consideration best way to **fill it is with median** value.
- **183 duplicated** data found after removing identifier, and I decide to drop it.
- After that, several **new features are extracted** from existing ones. The extracted columns are then dropped to avoid redundancy.
- **Manual label encoding** performed to ensure the values are ordered correctly
- Data scaling performed only for numerical datatype using **MinMaxScaler**

Number of Cluster

Silhouette Score & Inertia vs Number of Clusters

- For segmenting customer, there is a method called RFM Analysis, for you want to know deeply about RFM can read this reference : https://www.barilliance.com/rfm-analysis/#:~:text=RFM%20analysis%20is%20a%20data,much%20they've%20spent%20overall .
- * Recency   : Date of Last of Purchases `Recency`
- * Frequency   : Total Number of Orders `Total_Purchases`
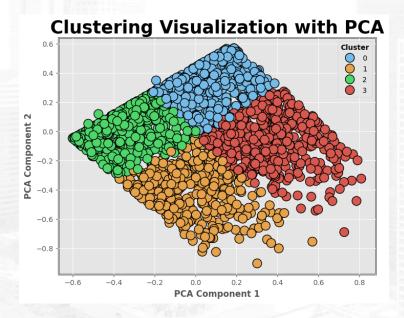- * Monetization : Total order value `Total_Amount`
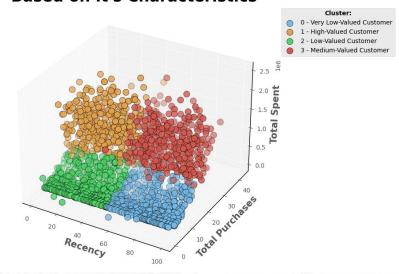- * Loyalty   : Total campaign accepted `Total_Campaigns_Accepted`
- Considering both the silhouette score and inertia score, we can observe that the elbow point, as well as the peak of the inertia score, occurs at **cluster number 4.**
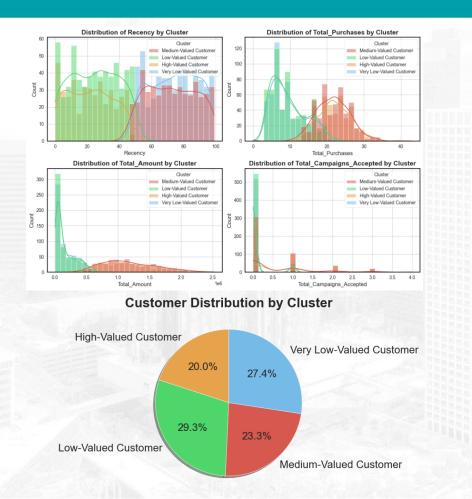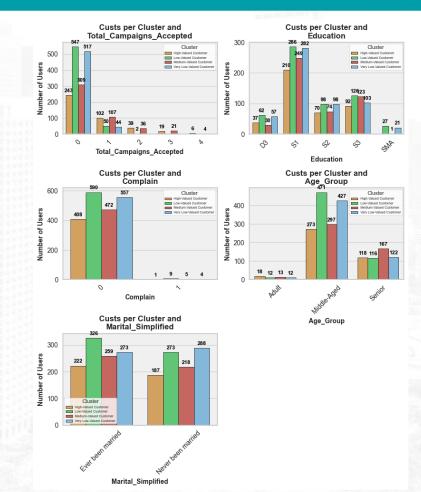
Customer Personality Cluster

Clustering Visualization with PCA

3-D Visualization of Customer Clusters Based on it's Characteristics

Customer Personality Analysis for Each Cluster

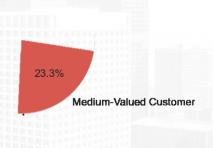**High-Valued Customer**

20.0%

Customers in this group have a low average recency (22 days), a high average total purchase (22 times), and a high average total amount (1.20 million Rupiah). 20% of our customers fall into this category.

Medium-Valued Customer

23.3%

Customers on this group have high average recency (72 days), high average total purchase (22 times) and high average total amount (1,18 Million Rupiah).There 24.09% of our customer fall into this category. There are 309 customer never accept our campaign, 107 accepted it once, 36 accepted it twice, 21 accepted it three times and 4 accepted it four times.

Low-valued Customer

29.3%

Customers on this group have low average recency (24 days), low average total purchase (10 times) and low average total amount (161K Rupiah).There 29.03% of our customer fall into this category. There are 547 customer never accept our campaign, 50 accepted it once and 2 accepted it twice.

Very Low-Valued Customer

27.4%

Customers on this group have high average recency (74 days), low average total purchase (10 times) and low average total amount (153K Rupiah).There 28,44% of our customer fall into this category. There are 517 customer never accept our campaign and 44 accepted it once.

Business Recommendation

Segment	Action and Implementation	Segment	Action and Implementation
High-Valued Customer	• **Loyalty Programs:** Offer exclusive rewards or VIP membership to retain them and increase lifetime value. • **Upsell & Cross-sell:** Promote premium or complementary products to increase basket size. • **Personalized Communication:** Keep them engaged with early access to new launches and personalized offers. • **Referral Incentives:** Encourage referrals through incentives since they're loyal and active.	**Low-Valued Customer**	• **Incentivize First Bigger Purchase:** Offer discounts or free shipping on higher-value items. • **Educate & Nurture:** Use content (e.g., how-tos, reviews) to build product value and trust. • **Bundle Offers:** Create appealing product bundles to boost average order value. • **Analyze Barriers:** Investigate what's preventing higher spending—price, product relevance, etc.
Medium-Valued Customer	• **Re-engagement Campaigns:** Use targeted emails or limited-time offers to bring them back. • **A/B Test Campaign Formats:** Since campaign response is low, experiment with new messaging, timing, or incentives. • **Segment by Past Purchases:** Tailor recommendations based on what they used to buy. • **Urgency & Scarcity:** Promote FOMO-driven campaigns ("Only 2 days left!") to rekindle interest.	**Vary Low-Valued Customer**	• **Reactivation Campaigns:** Offer a "We Miss You" deal with strong incentives (e.g., 50% off, free gift). • **Exit Surveys:** If they churn, ask why — to gather insights for improving products or communication. • **Exclude from Costly Campaigns:** Avoid targeting this segment with expensive marketing unless re-engagement ROI is proven. • **Retarget via Low-Cost Channels:** Use social media or SMS instead of email blasts or paid ads.

Business Simulation

Rakamin Academy

Simulation Scenario: Targeted Revenue Growth Plan

Assumptions:

Total Customers: 1,000

📝 **Segment Distribution:**

- High-Valued   : 20% (200 customers)
- Medium-Valued   : 24.09% (240 customers)
- Low-Valued   : 29.03% (290 customers)
- Very Low-Valued   : 28.44% (280 customers)

💰 **Current Revenue Per Segment:**

- High-Valued   : 200 × Rp 1.20M = Rp 240M
- Medium-Valued   : 240 × Rp 1.18M = Rp 283.2M
- Low-Valued   : 290 × Rp 161K   = Rp 46.69M
- Very Low-Valued   : 280 × Rp 153K   = Rp 42.84M
- 📌 Total Baseline Revenue Before ML: **Rp 612.73M**

Segment	Strategy	Targeted Uplift	Projected Gain
High-Valued Customer	Loyalty perks & premium upsell	+10% spend/customer	+Rp 24M
Medium-Valued Customer	Win-back campaign (15% reactivation)	36 customers × Rp 1.18M	+Rp 42.5M
Low-Valued Customer	Bundled discounts (Increase avg. spend by Rp 50K)	290 customers × Rp 50K	+Rp 14.5M
Very Low-Valued Customer	Final reactivation offer (5% reactivation)	14 customers × Rp 153K	+Rp 2.14M
	💡 **Final Business Impact After ML** Projected New Total Revenue : Estimated Campaign Costs : Total Additional Revenue :		Rp 695.87M Rp 15M **Rp 83.14M (+11.1% increase)**