

Customer Personality Analysis and Predictive Modeling for Marketing Campaign Optimization

Widyadari Dhiya Parahita



[Link Code](#)



[Deploy](#)



Widyadari Dhiya Parahita

Working

Fresh Graduate

Education



Universitas Indonesia

Statistics

Aug 2020 - Feb 2025



DQ Lab

Jul 2021 - Aug 2021



Rakamin Academy

Data Science: ML Bootcamp

Mar 2024 - Aug 2024

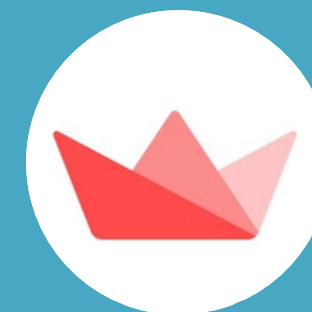


Dibimbing.id

AI/ML Bootcamp

Jun 2025 - Dec 2025

Tools



Abstract geometric shapes on the left side of the slide: a large black parallelogram, a medium yellow parallelogram, and a smaller orange parallelogram, all slanted to the right.

Problem Statement

Problem Statement

The company wants to identify customers who are likely to respond positively to marketing campaigns (response column = 1). However, the number of customers who respond is very small compared to those who do not respond, **resulting in class imbalance**.

In a business context, **failing to detect customers who are actually interested (False Negative)** will cause the company to lose conversion opportunities, reduce campaign effectiveness, and waste marketing budget allocation.

Therefore, the main objective of the model is not only to achieve high accuracy, but **to maximize Recall for the positive class (response = 1)** — that is, to ensure that as many genuinely interested customers as possible are detected by the model.

Metrics: Recall

$$\frac{TP}{TP + FN}$$

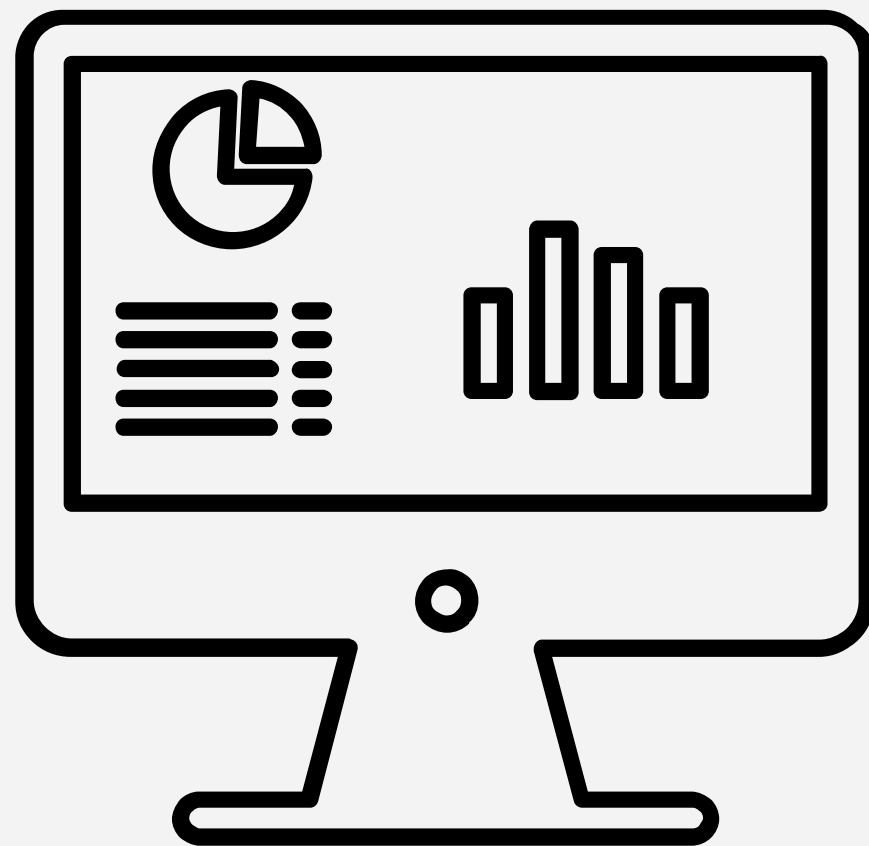
Abstract geometric shapes in black, light orange, and dark orange on the left side of the slide.

Data Understanding

About Dataset



2240 Sample



Customer Personality Dataset

Source: [Kaggle](https://www.kaggle.com/datasets/ibimbing/customer-personality-dataset)



25 Columns:

9 Categories, 15 Numeric

1. Income
2. Kidhome
3. Teenhome
4. Recency
5. NumWebVisitsMonth
6. NumDealsPurchases
7. NumWebPurchases
8. NumCatalogPurchases
9. NumStorePurchases

10. MntWines
11. MnstFruits
12. MntMeatProducts
13. MntFishProducts
14. MntsSweetProducts
15. MntGoldProds

1. acceptedcmp1
2. acceptedcmp2
3. acceptedcmp3
4. acceptedcmp4
5. acceptedcmp5
6. Education
7. Marital_Status
8. Complain
9. Response (Target)

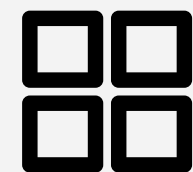
monetary

Data Quality



Data Imbalance

Response	Value Count
0	1906
1	334



Chi-Square Test

p-value Complain > 0.05



Missing Values

Income: 24 samples (0,01%)



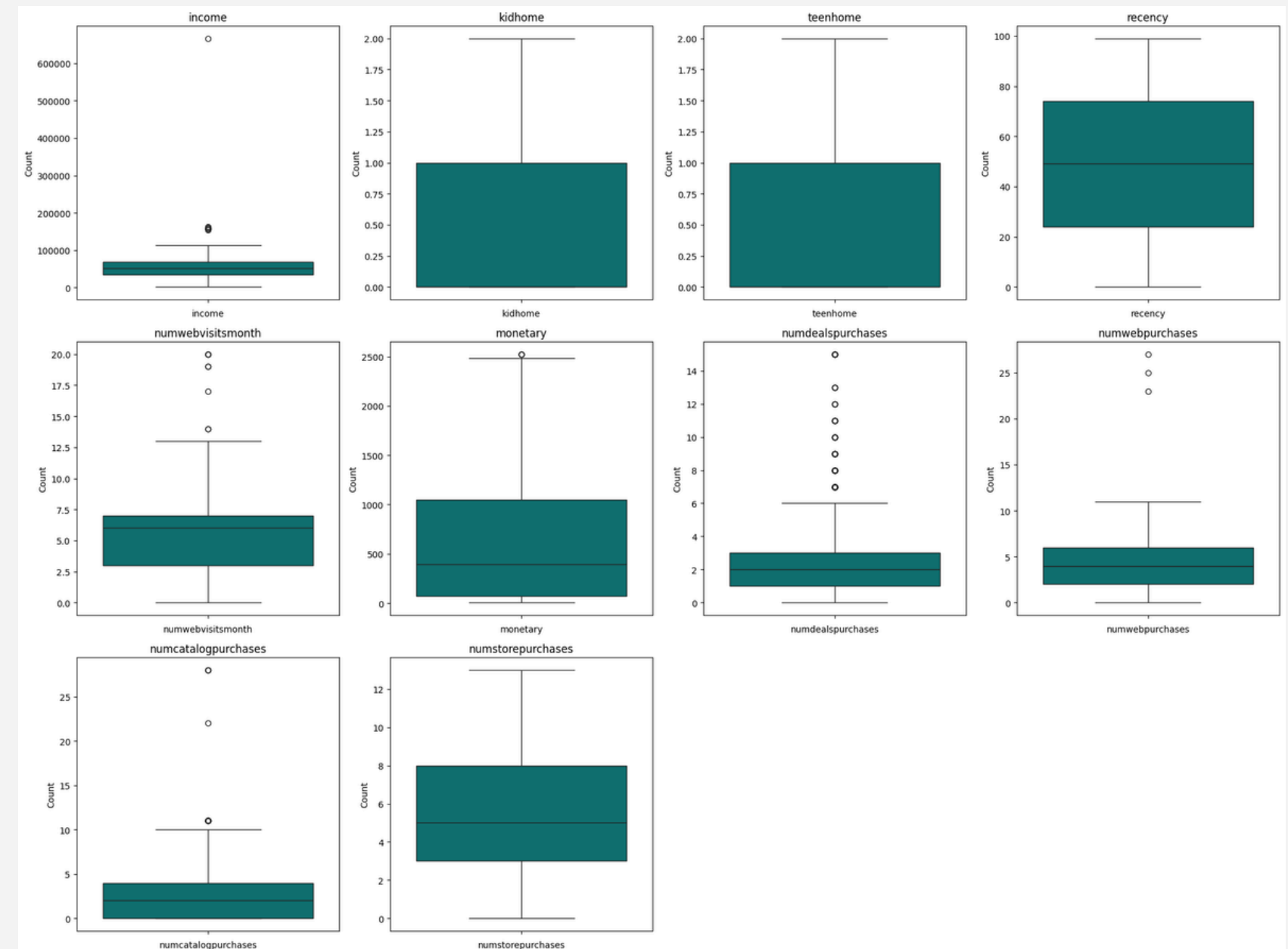
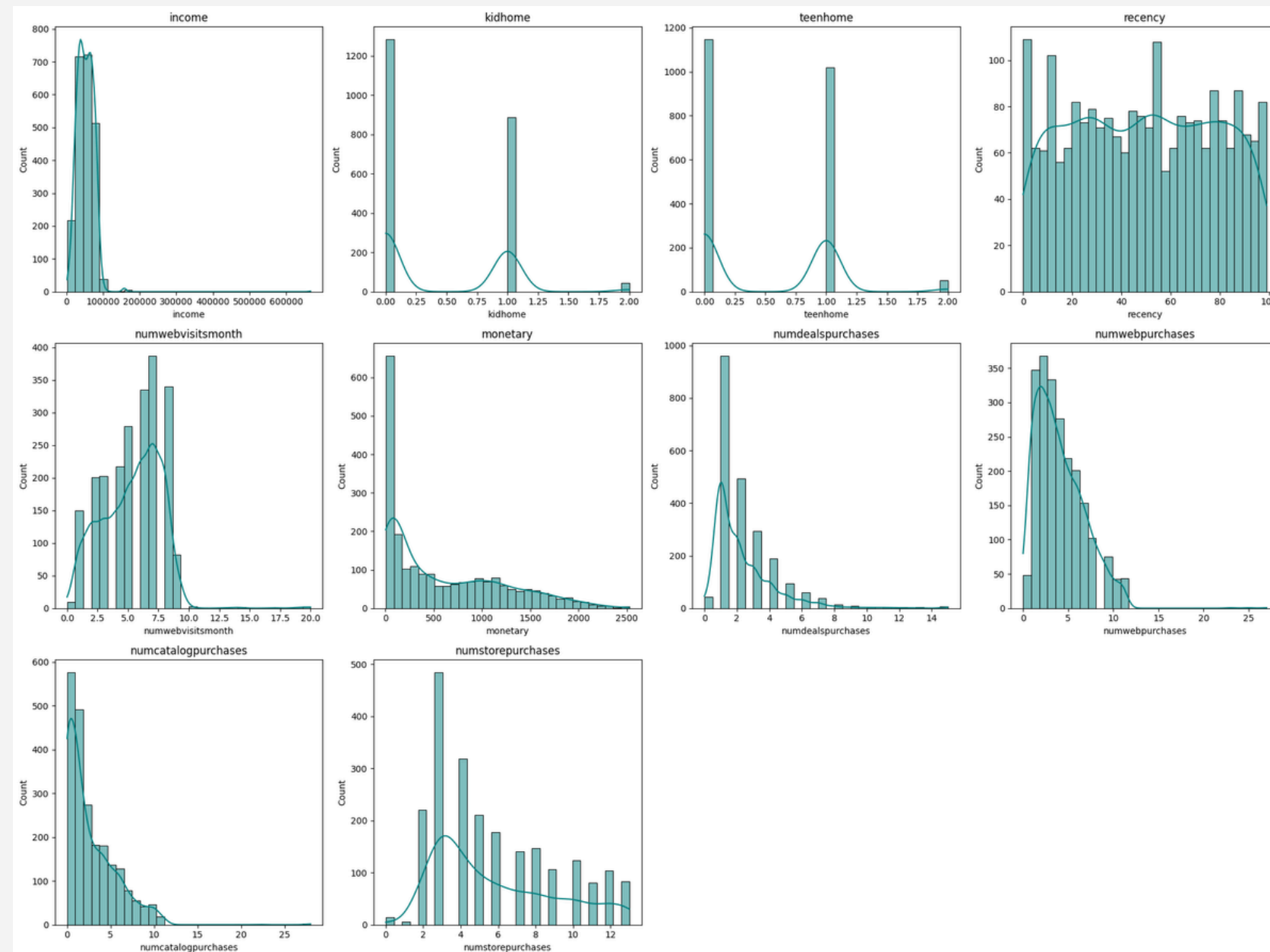
Multicollinearity

VIF < 10

feature	VIF
income	8.557253
kidhome	2.839022
teenhome	2.487744
recency	3.503716
numwebvisitsmonth	6.152165
monetary	8.735268
numdealspurchases	3.834974
numwebpurchases	5.975202
numcatalogpurchases	5.102932
numstorepurchases	7.938349

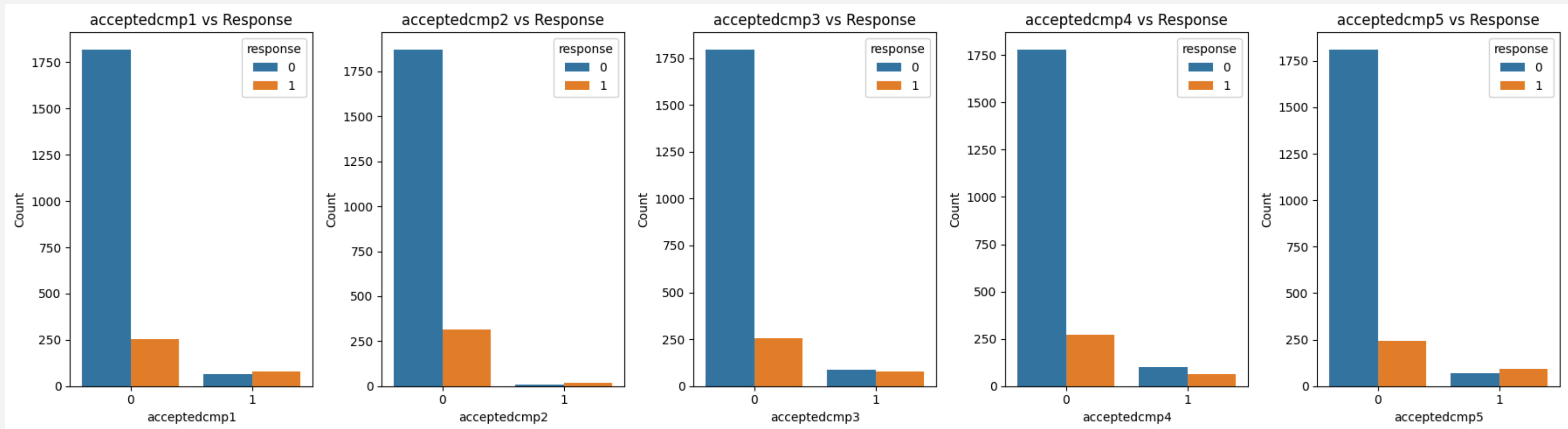
Data Quality

Min-Max Scaler



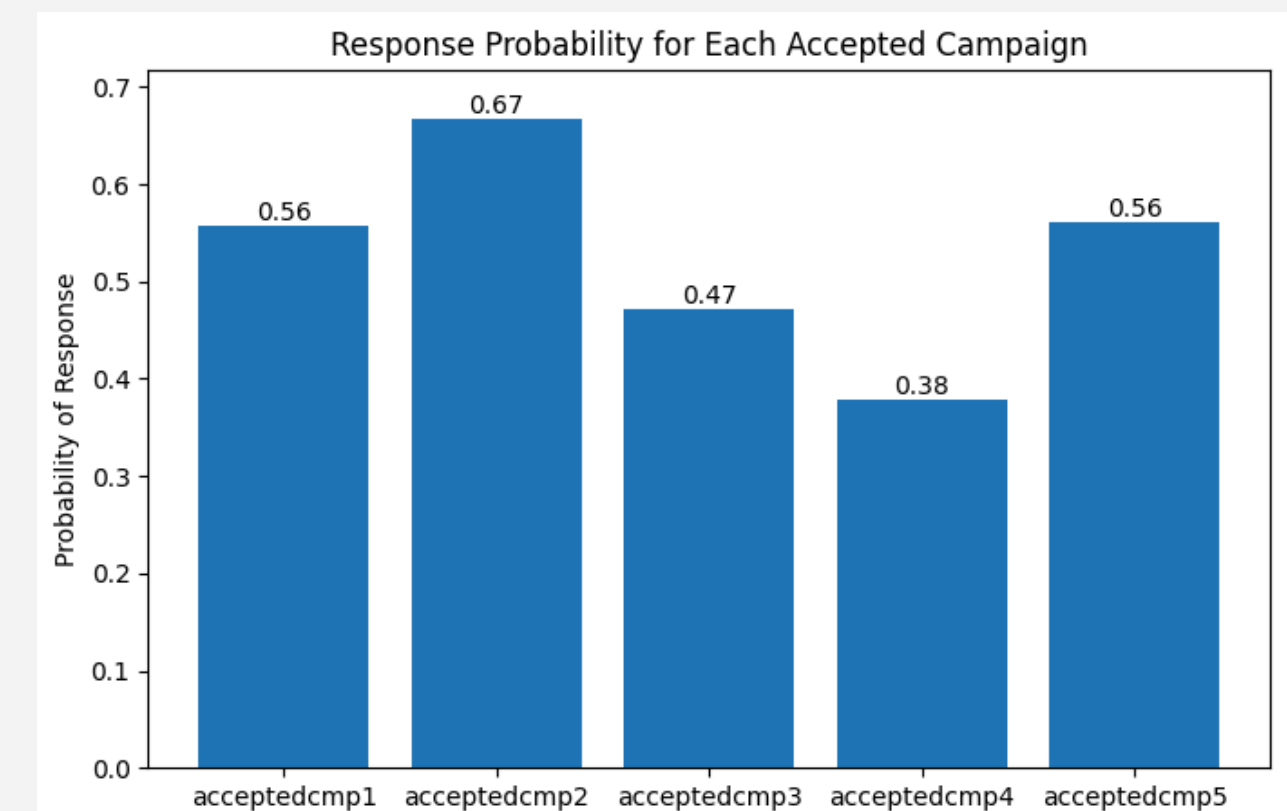
A series of overlapping geometric shapes on the left side of the slide: a large black parallelogram, a medium yellow parallelogram, and a smaller orange parallelogram, all slanted to the right.

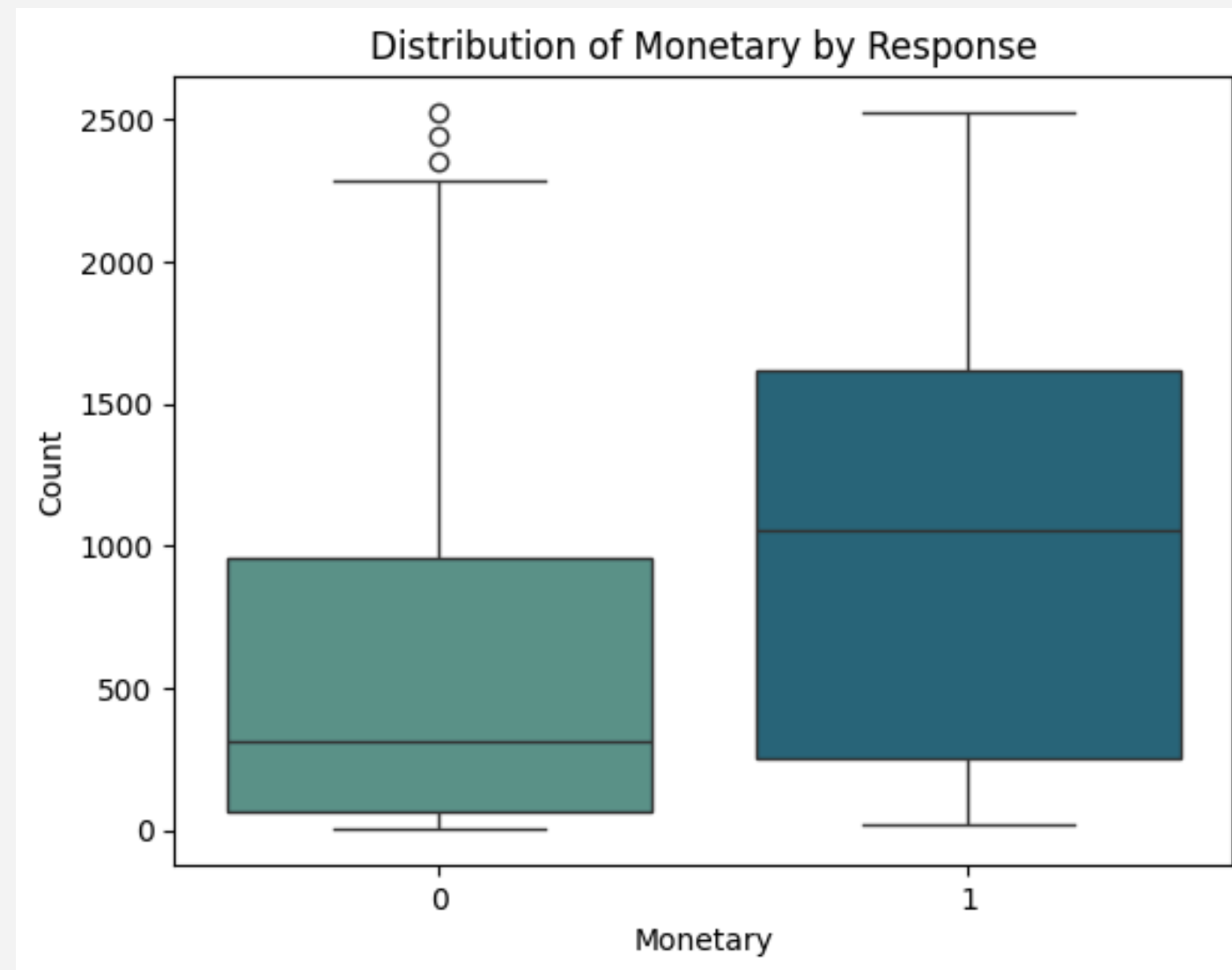
Exploratory Data Analysis



These customers:

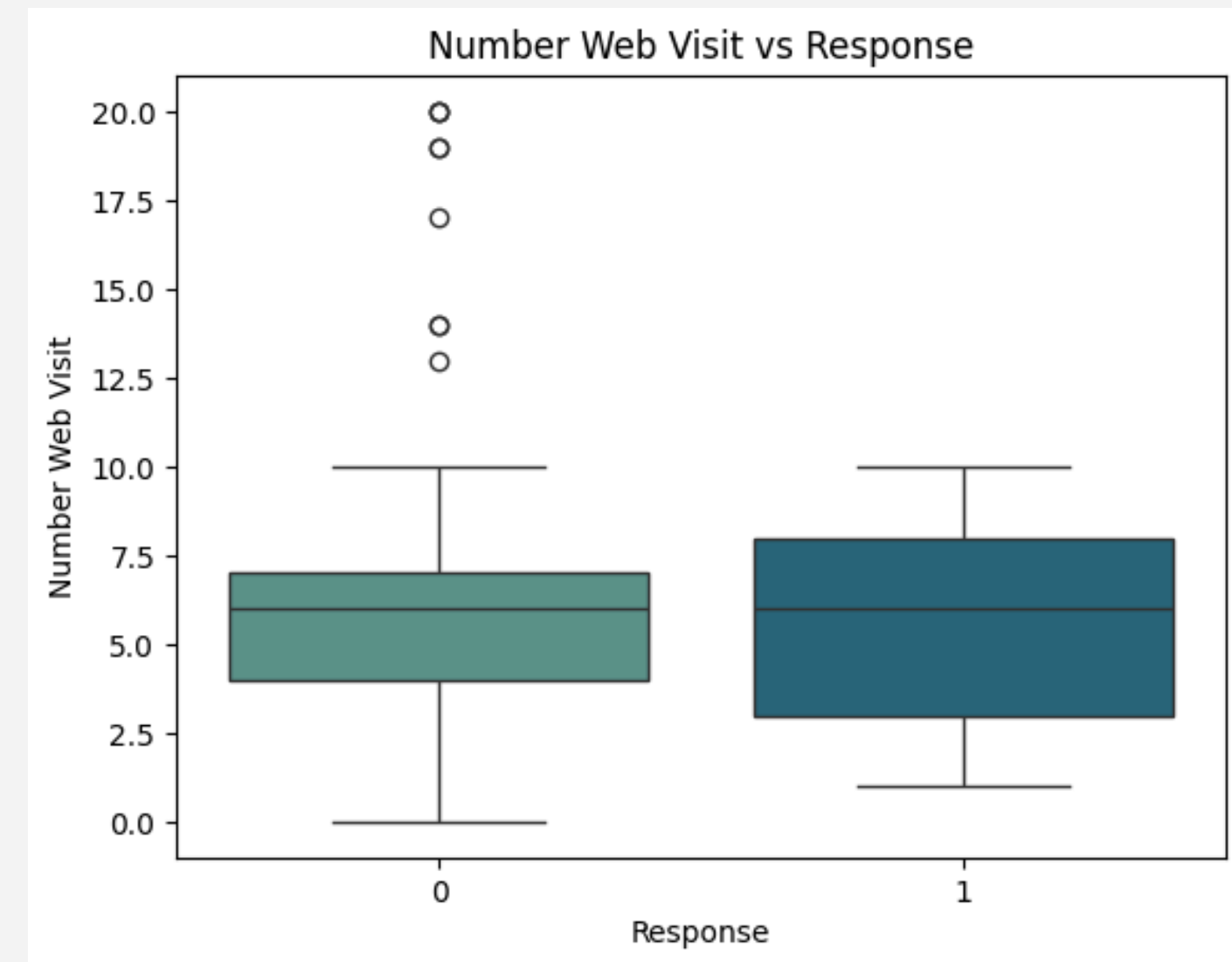
- Show consistent engagement behavior
- Are more open to marketing communications
- Have a higher probability of responding positively to new offers
- acceptedcmp2 show the highest response rate among customers who previously engaged with them.



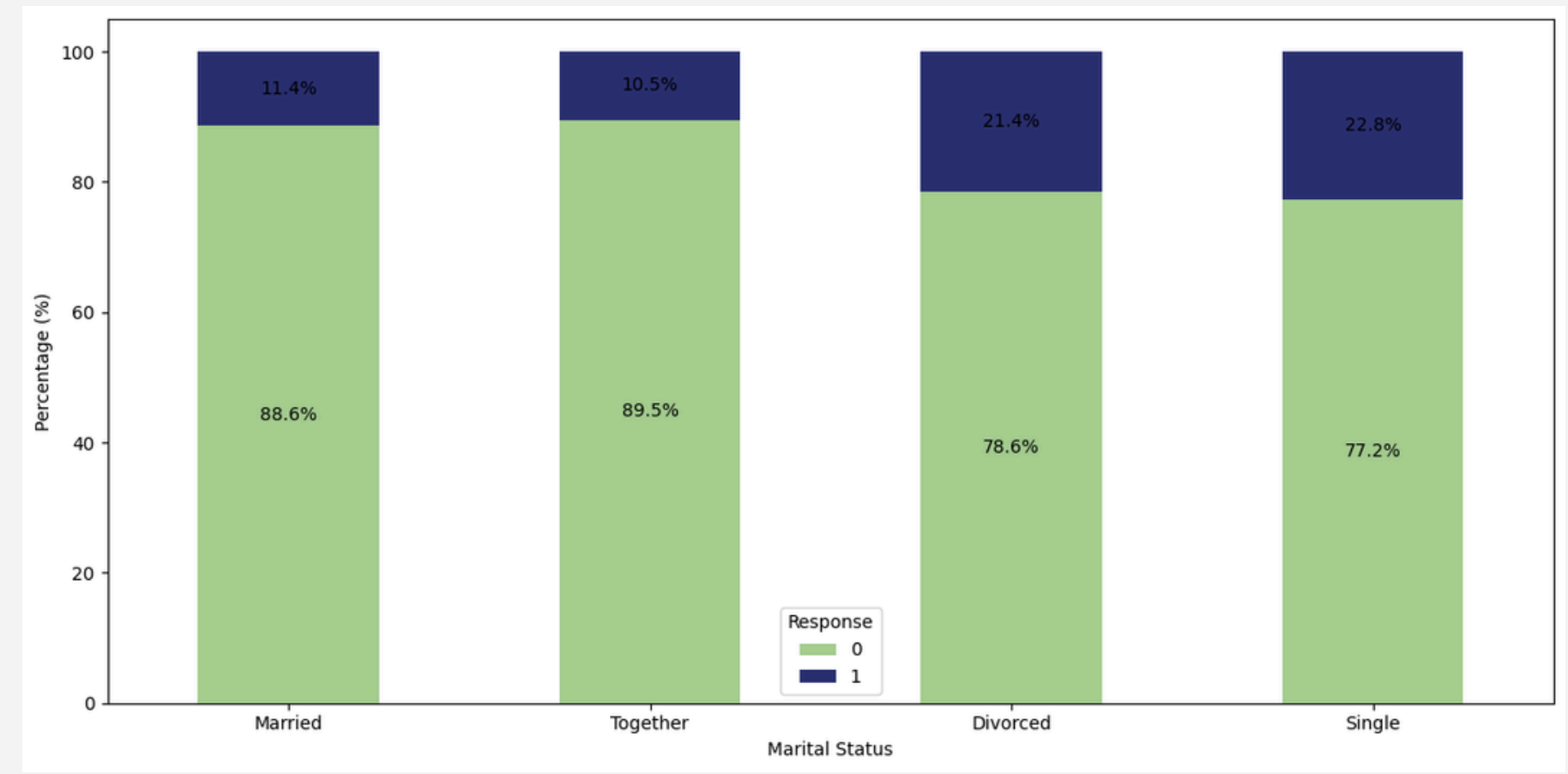
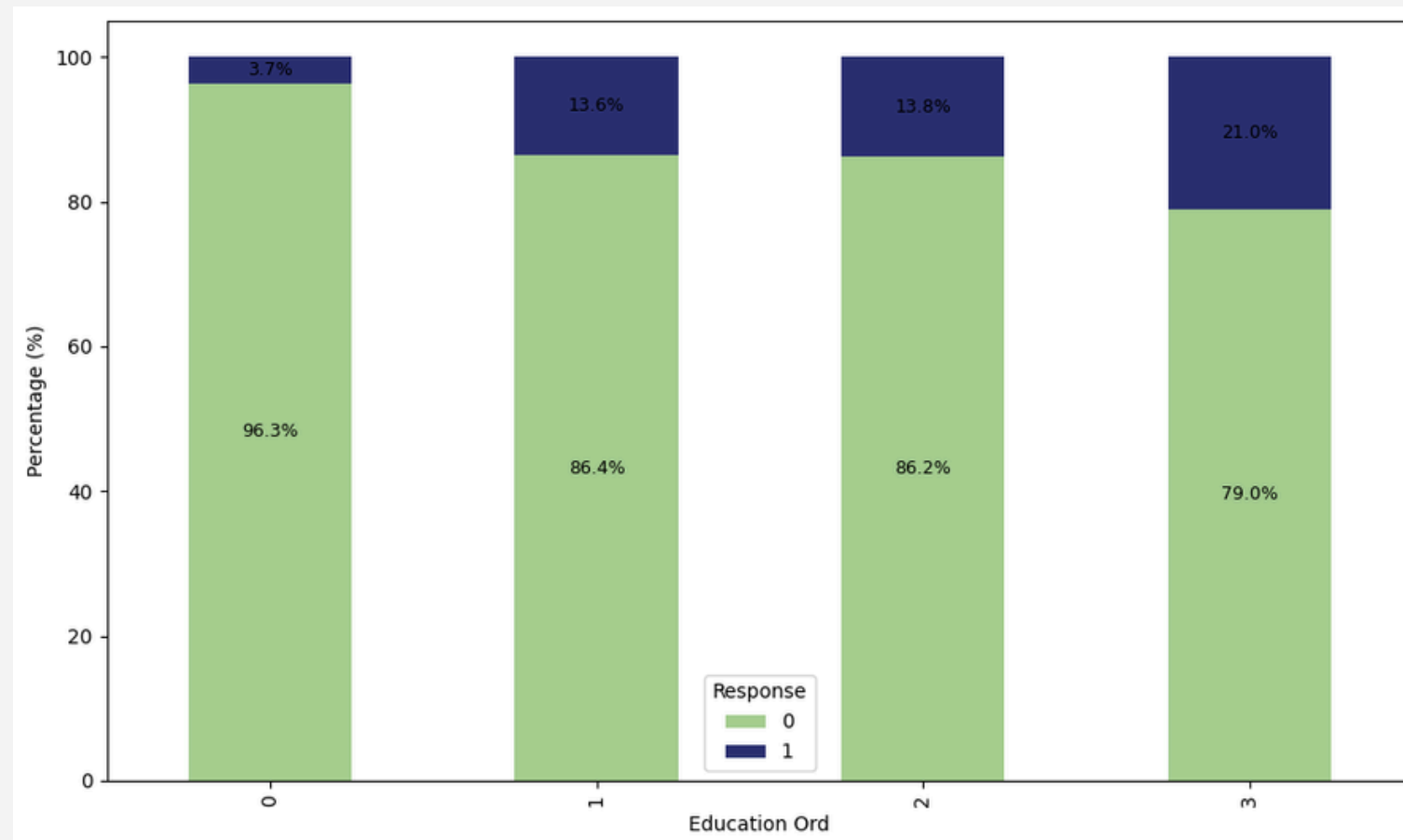


Customers who respond to campaigns are usually high-value customers.

- They are accustomed to making large purchases
- They are more sensitive to promotions or additional offers
- They have the potential to provide high ROI if targeted correctly



- Digital engagement (web activity) is positively correlated with campaign response.
- Customers who actively visit websites are usually more aware of brands and more easily influenced by campaigns.




The most responsive customer groups are:

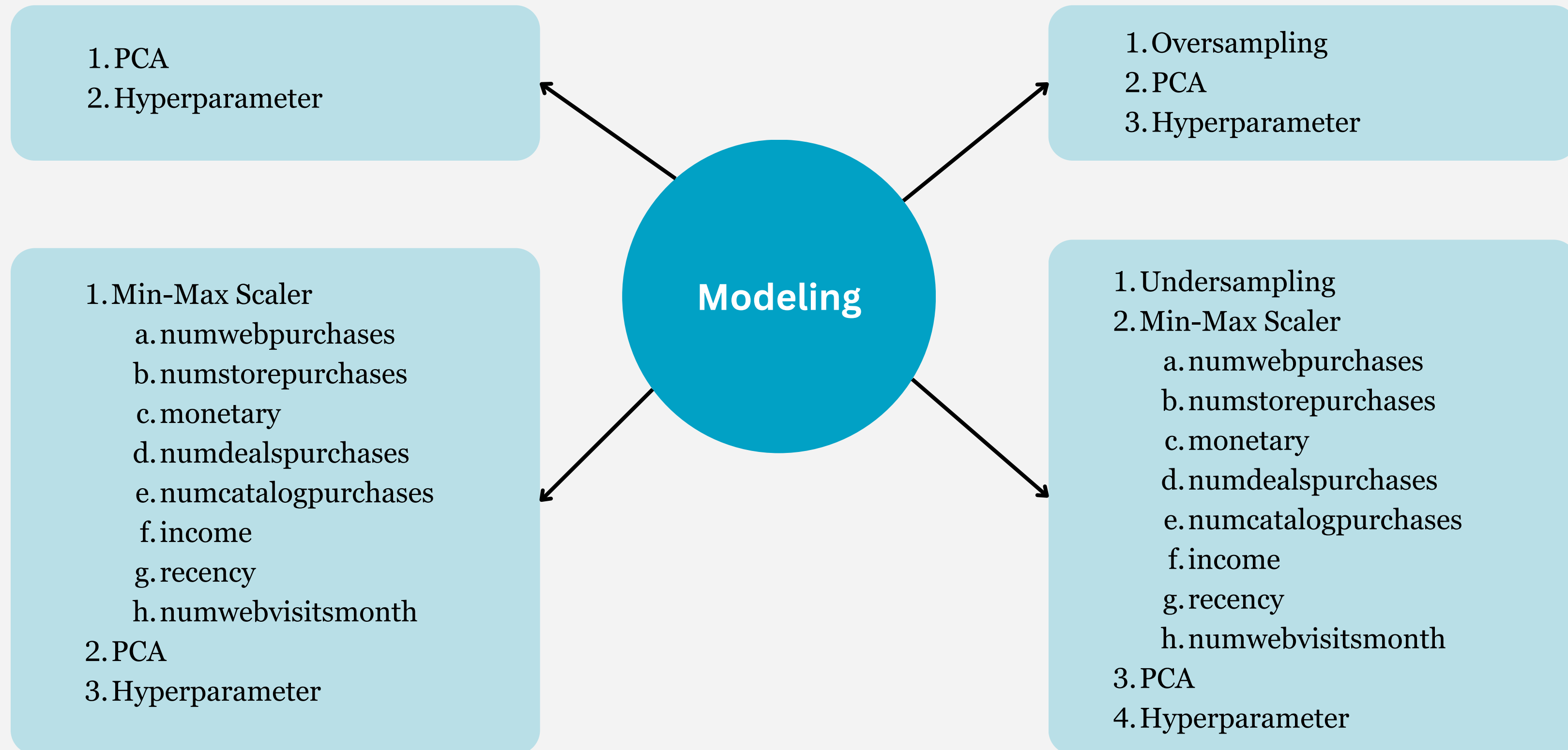
1. Single or Divorced

2. Highest education level (Education Ord = 3) These groups share characteristics:

- More autonomy in decision-making
- Likely higher purchasing power (for higher education)
- Possibly more active in digital engagement patterns

A series of overlapping geometric shapes on the left side of the slide: a large black parallelogram, a light orange parallelogram, and a darker orange parallelogram, all slanted to the right.

Model Selection and Training



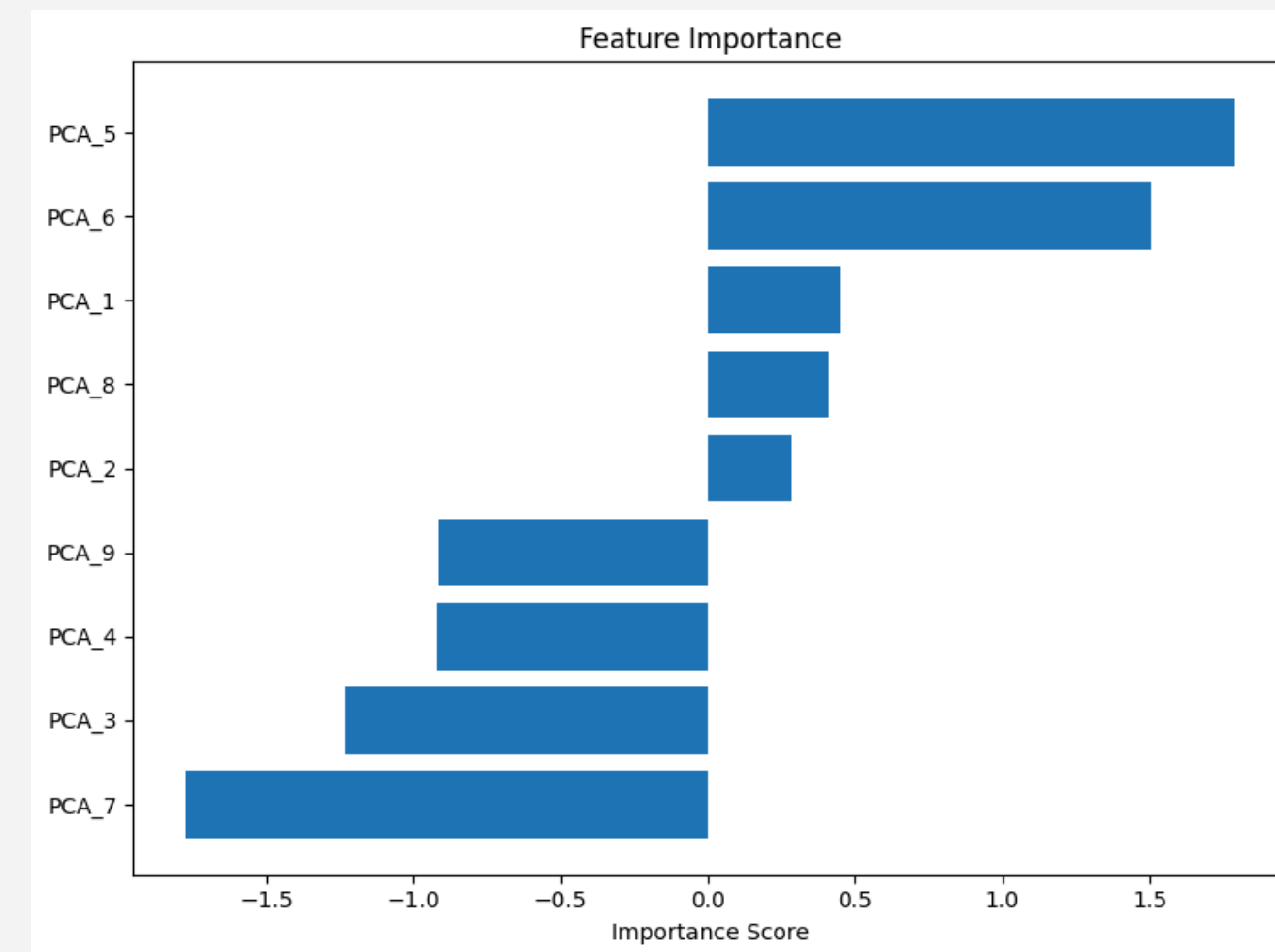
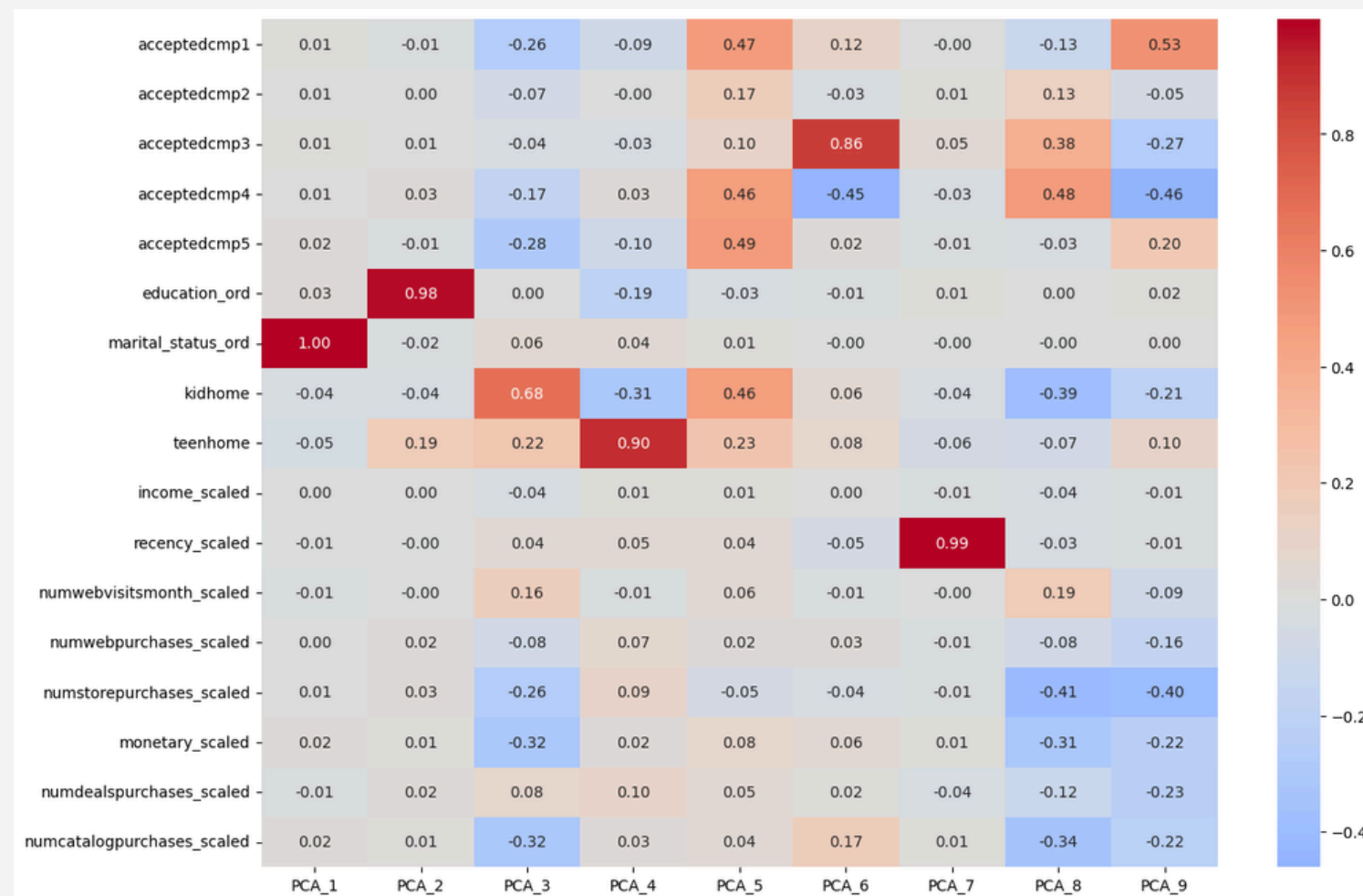
A series of overlapping geometric shapes on the left side of the slide: a large black parallelogram, a light orange parallelogram overlapping its bottom-right corner, and a darker orange parallelogram overlapping the bottom-right corner of the light orange one.

Model Evaluation

Evaluation Metrics: Recall

$$\frac{TP}{TP + FN}$$

Model	Train Accuracy	Test Accuracy	Train Precision	Test Precision	Train Recall	Test Recall	Train F1	Test F1
Logistic Regression	0.783404	0.783459	0.657795	0.376289	0.729958	0.760417	0.692000	0.503448
KNN	0.825598	0.843609	0.815642	0.459184	0.616034	0.468750	0.701923	0.463918
Decision Tree	0.997187	0.774436	1.000000	0.359375	0.991561	0.718750	0.995763	0.479167
Random Forest	0.997187	0.845113	0.991632	0.471545	1.000000	0.604167	0.995798	0.529680
Gradient Boosting	0.939522	0.845113	0.975490	0.472868	0.839662	0.635417	0.902494	0.542222
XGBoost	0.997187	0.822556	0.991632	0.432927	1.000000	0.739583	0.995798	0.546154

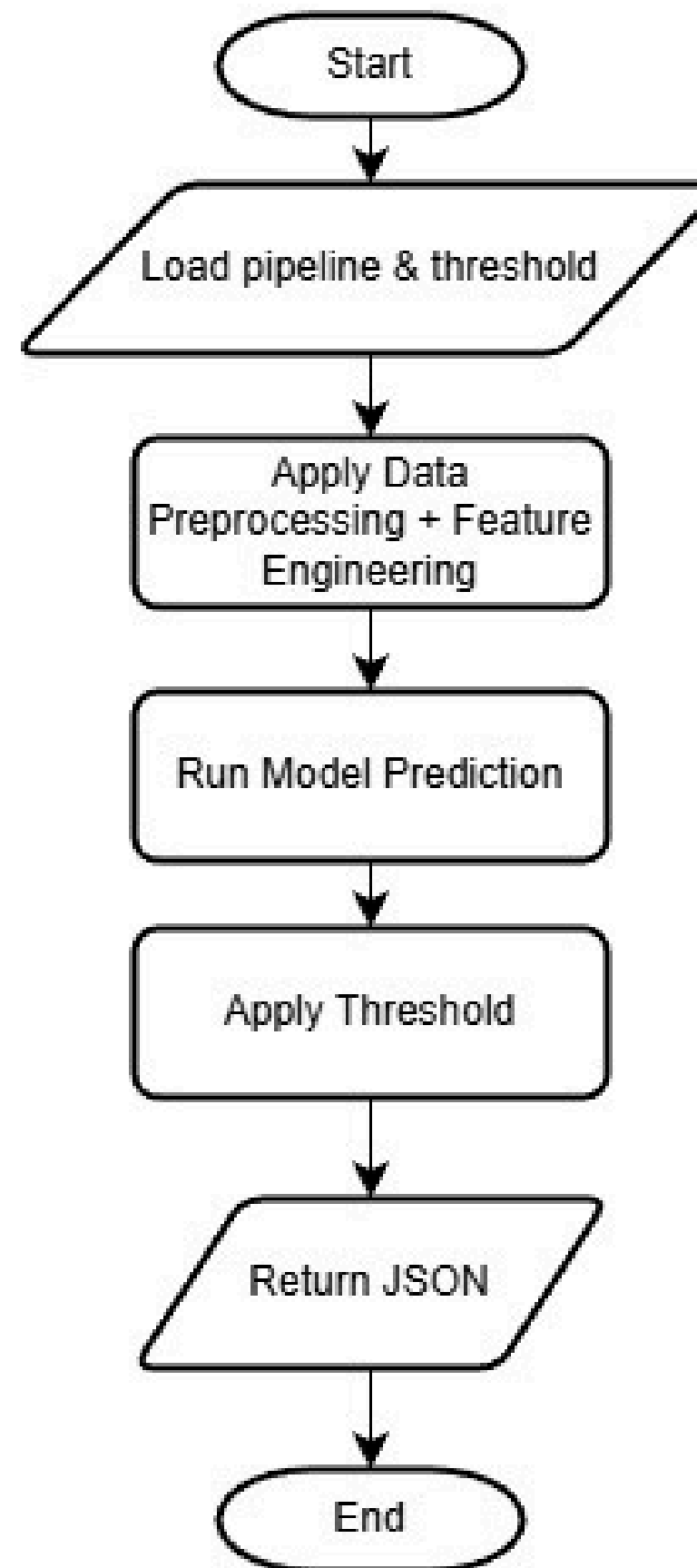
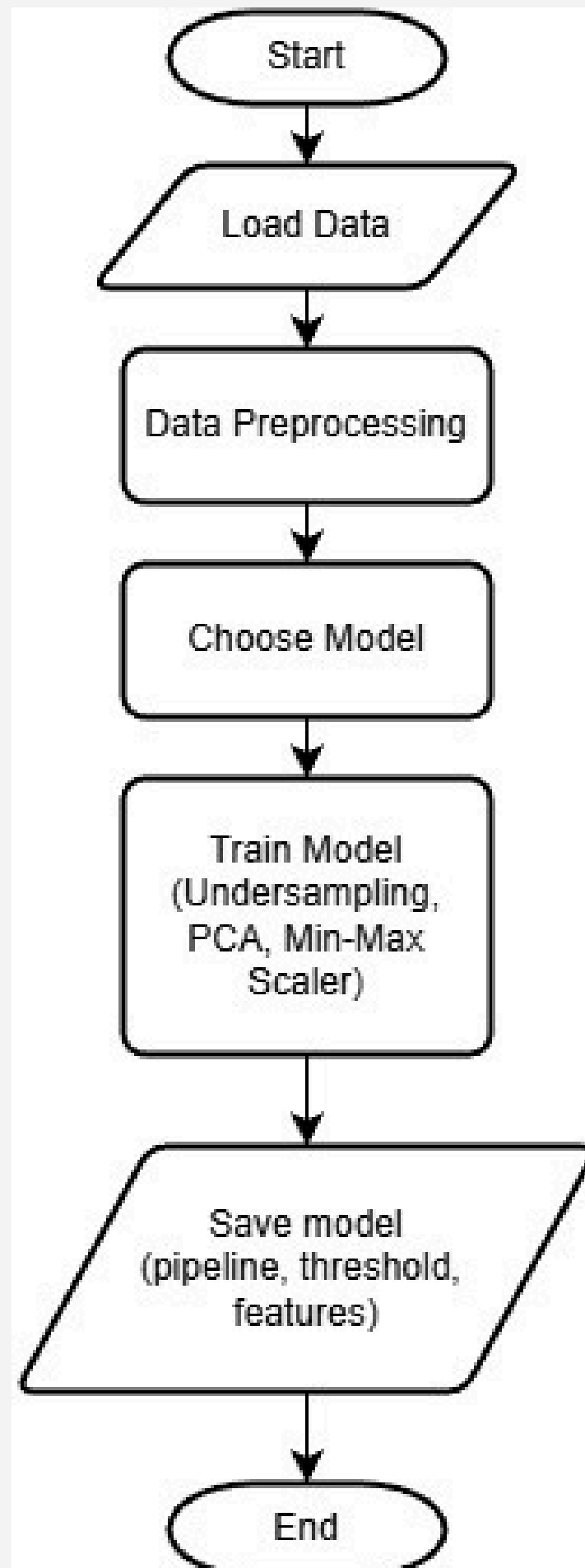


Insight:

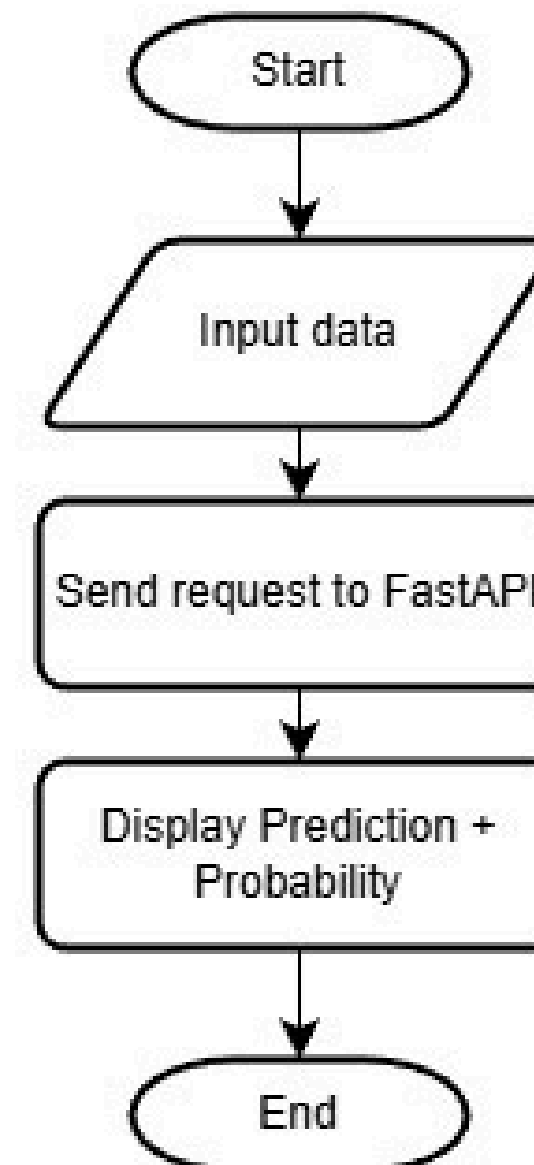
- The model is mainly influenced by PCA 5, PCA 6, and PCA 7.
- PCA 7 has a large negative feature importance, **customers with lower recency** (more recent purchases) **have a higher probability of responding to the campaign**. This aligns with marketing logic: **recently active customers are easier to convert**.
- PCA 5 represents **household-driven and campaign-engaged customers** (teenhome, kidhome, cmp4, cmp5).
- PCA 6 represents **campaign behavior patterns** (especially cmp3 and cmp4), **plus moderate influence from catalog usage and family structure**.

Abstract geometric shapes in black, light orange, and dark orange on the left side of the slide.

Deployment



FASTAPI



STREAMLIT

Decorative geometric shapes on the left side of the slide: a large black parallelogram, a light orange parallelogram, and a darker orange parallelogram, all slanted to the right.

Conclusion & Recommendation

Recommendation

1. Focus on customers with low recency, strong spending patterns, and prior campaign engagement to maximize ROI
2. Customer who previously accepted campaigns should be first in the targeting queue due to their proven responsiveness
3. Increase use of website reminders and personalized recommendations triggers to leverage high digital engagement behavior
4. Single, Divorced, and highly educated customers should receive tailored messaging, as they show higher openness to marketing offers

1. Use logistic regression as the production model
2. Lower the decision threshold (<0.5) if the business wants higher recall and broader customer capture.
3. Try another model

A large decorative graphic on the left side of the slide, consisting of a light blue circle containing a smaller orange circle, which in turn contains a smaller light blue circle, creating a nested circle effect.

Thank you



[Deployment | Streamlit](#)



[Widyadari Dhiya Parahita | LinkedIn](#)



[Widyadari Dhiya Parahita | Github Portfolio](#)



widyadari.dhiya@gmail.com



[Jakarta, Indonesia](#)



CERTIFICATE OF APPRECIATION

This certificate is proudly presented to

Widyadari Dhiya Parahita

in recognition of his outstanding record of contribution as the Best Final Project in
AI Machine Learning Bootcamp Batch 7

Presented this December of 2025

A handwritten signature in black ink, appearing to read 'Zaky Muhammad Syah', written over a faint ibimbing logo.

Zaky Muhammad Syah
Chief Executive Officer

A handwritten signature in black ink, appearing to read 'Tiara Sri Mulati', written over a faint ibimbing logo.

Tiara Sri Mulati
Product Operation Lead