BUSINESS STRATEGY

Data-driven Decision Making

Werner Laemlin & **Nathan Dubourg**

Github: https://github.com/NaDub/visu_2025/

Project Overview

A company wants to produce the highest profit for the next direct marketing campaign.

Context:

- A pilot campaign targeted 2,240 customers.
- Customers were randomly selected and contacted by phone.
- Campaign cost: **6,720MU**, Revenue: **3,674MU**
- Profit: **-3,046MU**, Profit/Customer: **-1.36MU**.
- Success rate: **15%**.

Objectives

Our Goals

- Develop an accurate predictive model.
- Understand key customer characteristics.
- Maximize the profitability of marketing campaigns.

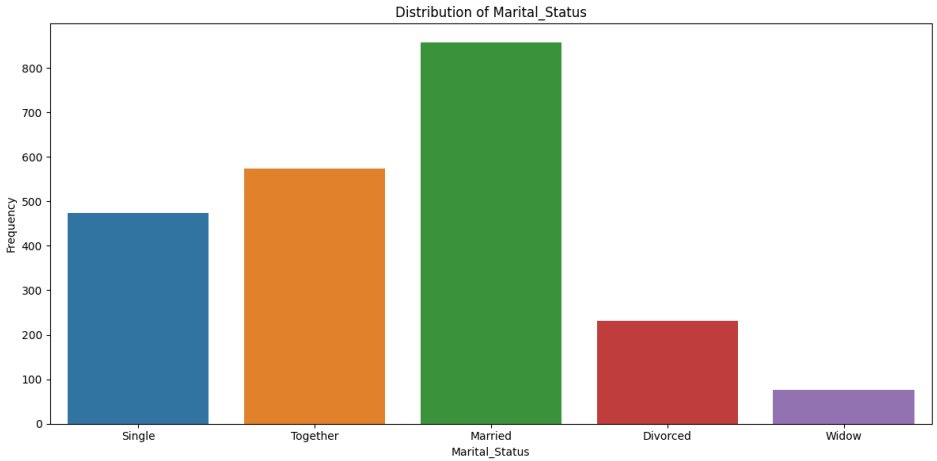
Data Exploration

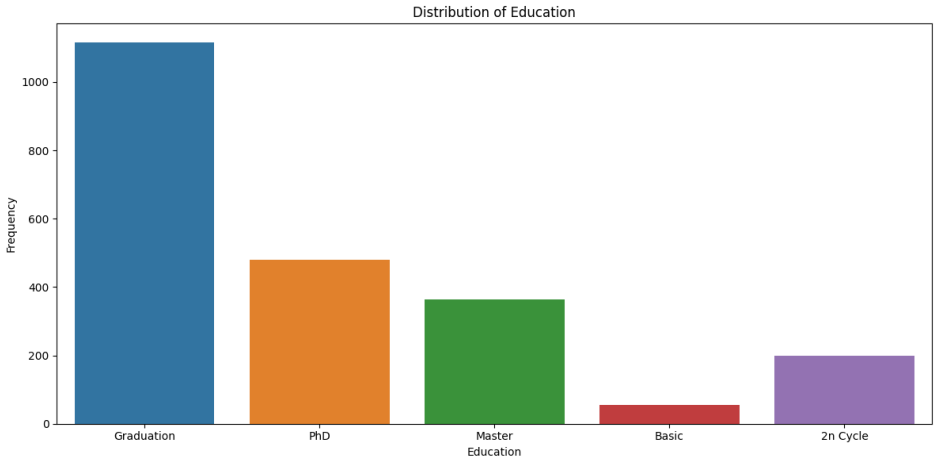
Unveiling Insights for Strategic Decisions

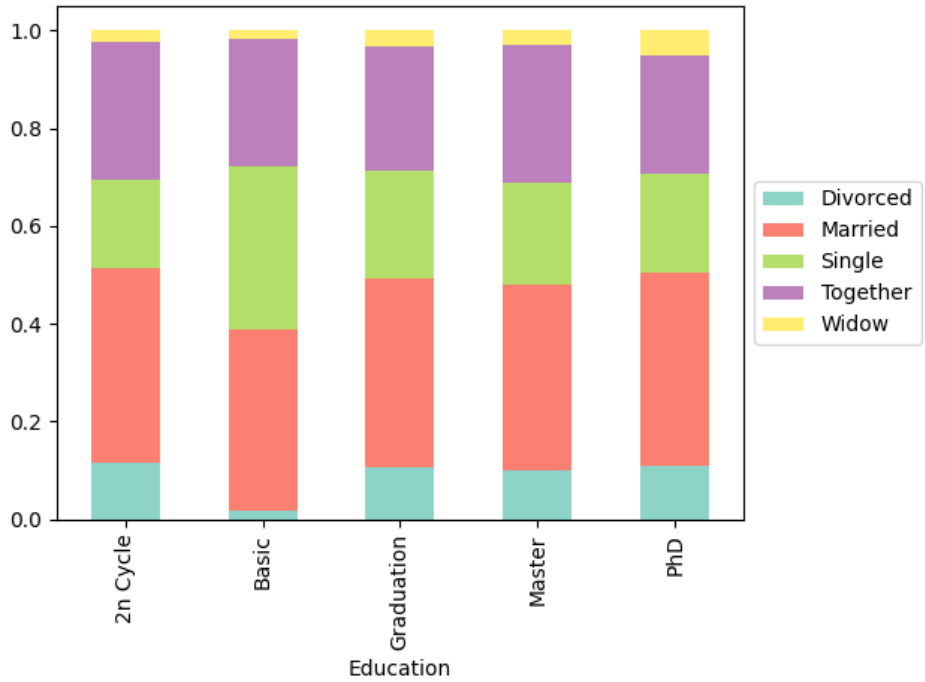
Explanation of Variables

- **AcceptedCmp1**: 1 if the customer accepted the offer in the 1st campaign, 0 otherwise.
- **Response (target)**: 1 if the customer accepted the offer in the last campaign, 0 otherwise.
- **Complain**: 1 if the customer complained in the last 2 years.
- **DtCustomer**: Date of the customer's enrolment with the company.
- **Education**: Customer's level of education.
- **Marital**: Customer's marital status.
- **Kidhome**: Number of small children in the customer's household.
- **Teenhome**: Number of teenagers in the customer's household.
- **Income**: Customer's yearly household income.
- **MntFishProducts**: Amount spent on fish products in the last 2 years.
- **NumDealsPurchases**: Number of purchases made with a discount.
- **Recency**: Number of days since the last purchase.
- **Customer_Age**: Seniority of the consumer.

Customer profiles



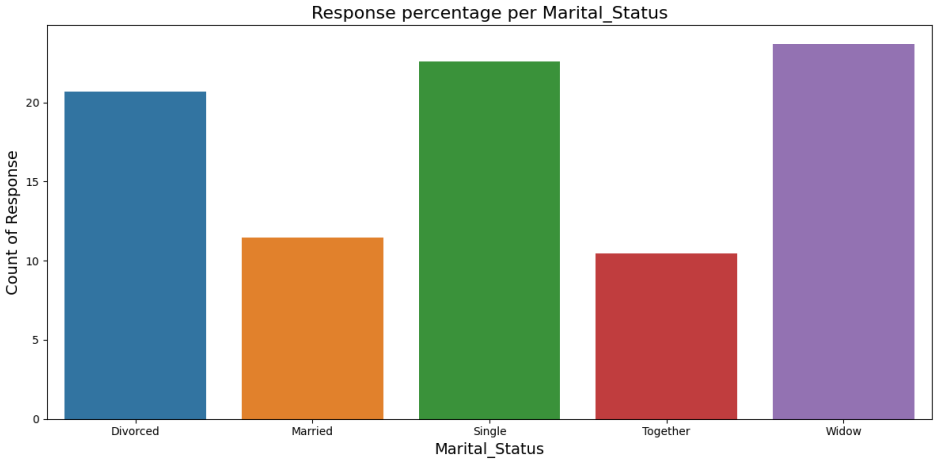


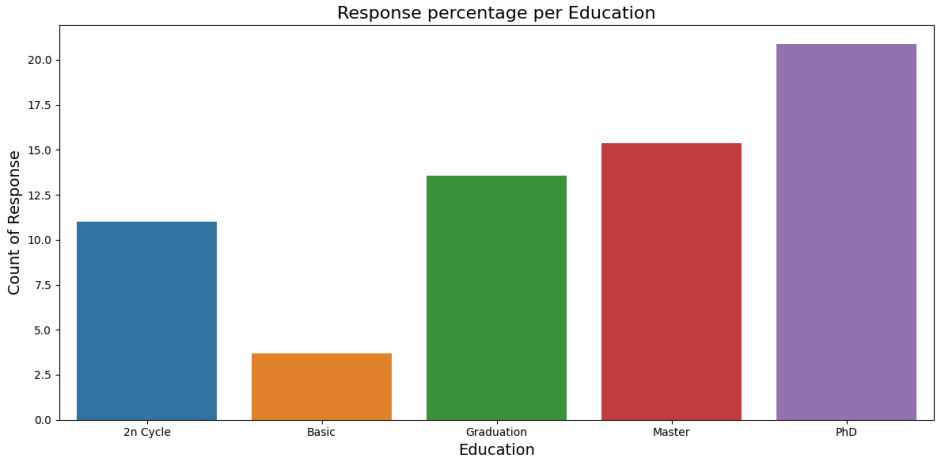


Customer Profile:

- Mostly married and highly educated.
- No significant correlation between these factors.

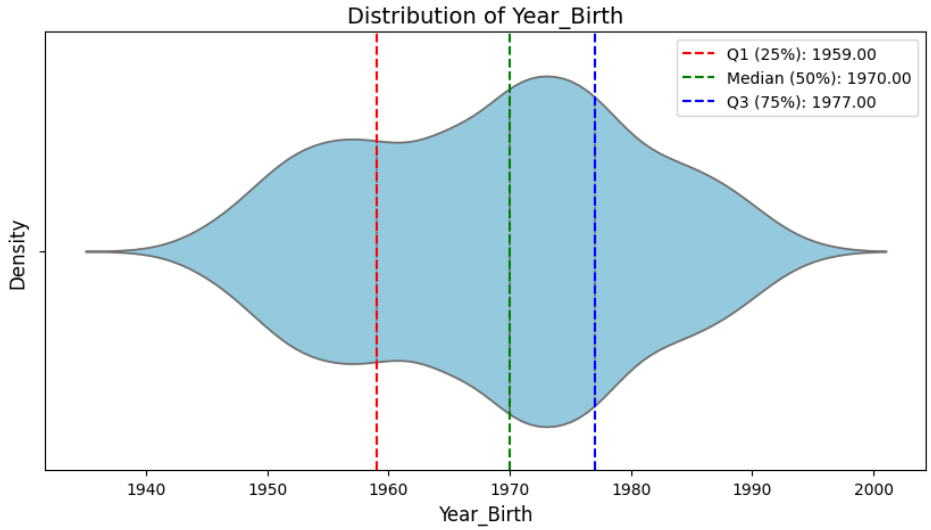
Customer profiles: Percentage of response

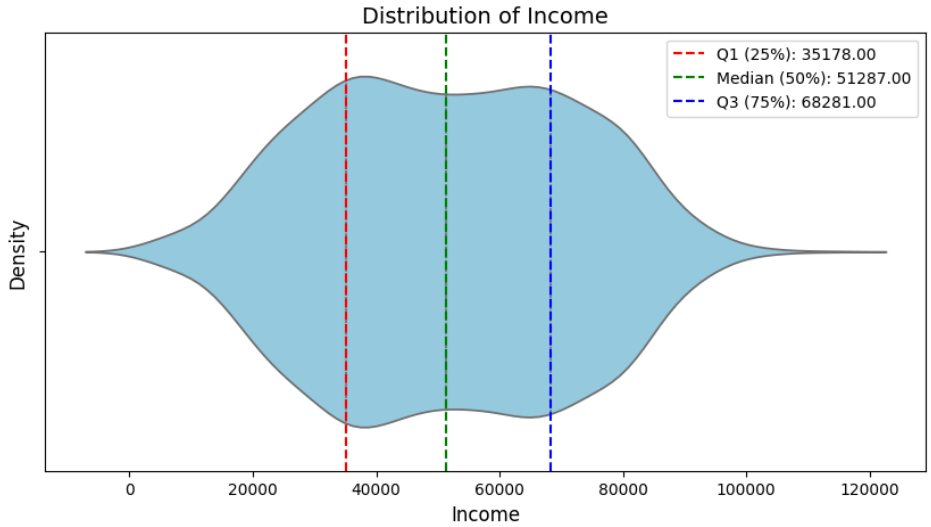




- The rate of positive responses in the last campaign appears to be higher if the consumer is not in a relationship.
- The rate of positive responses seems to increase with the level of education.
- ⚠ These relationships are not confirmed in previous campaigns.

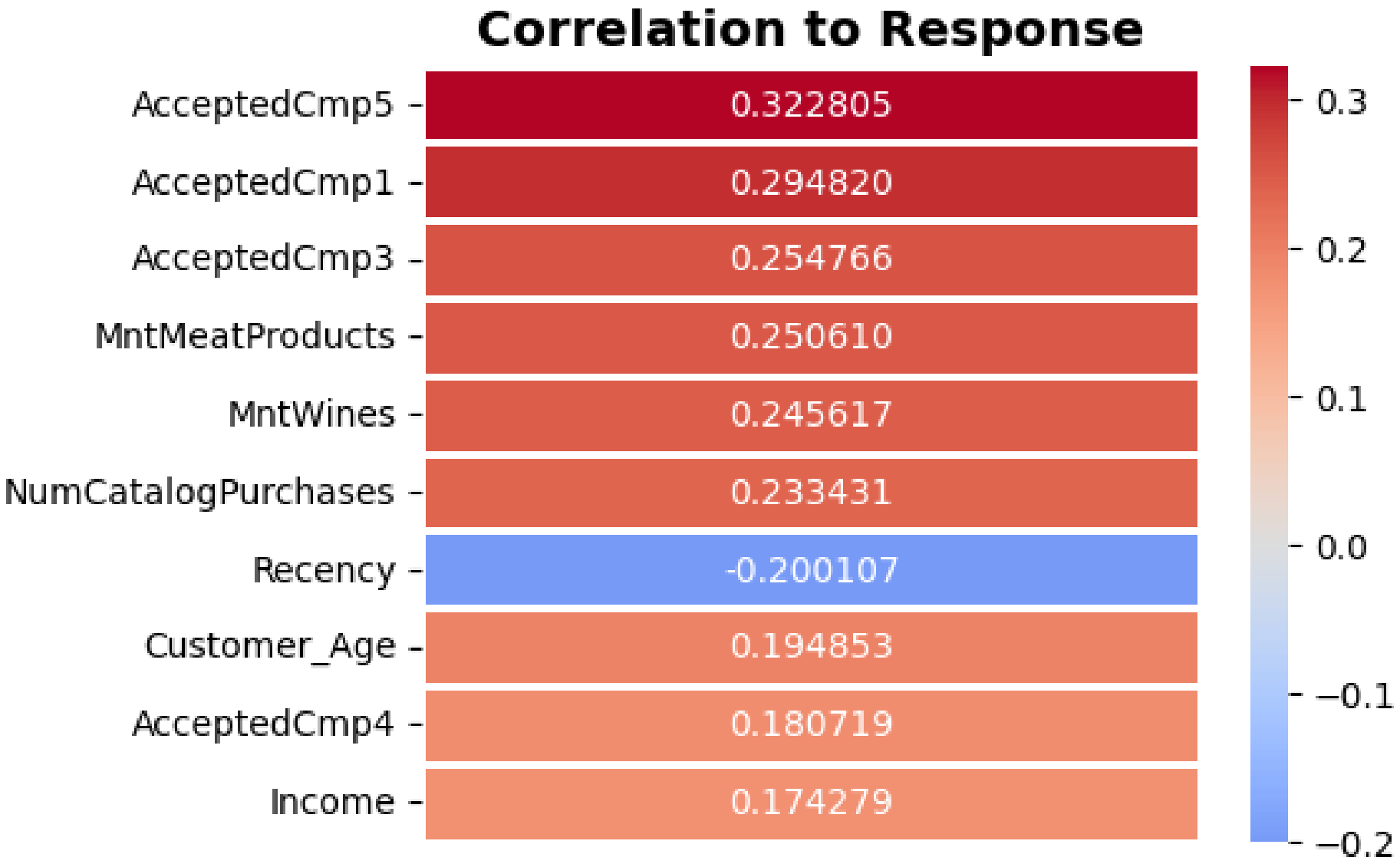
Customer profiles: Distributions of Year birth and Income





- A significant share of consumers were born between 1970 and 1977.
- The income distribution is close to a normal distribution.

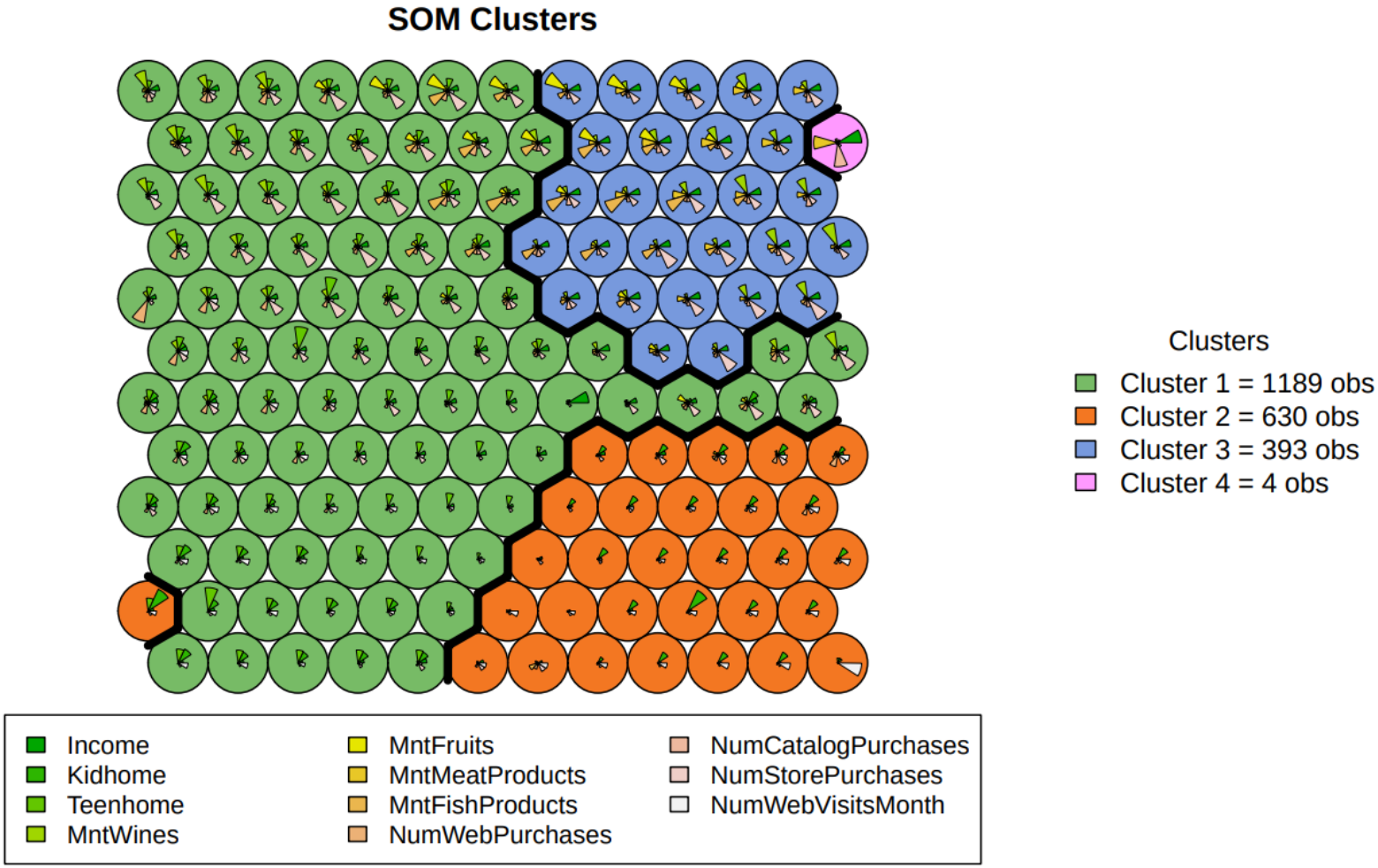
Key Correlations



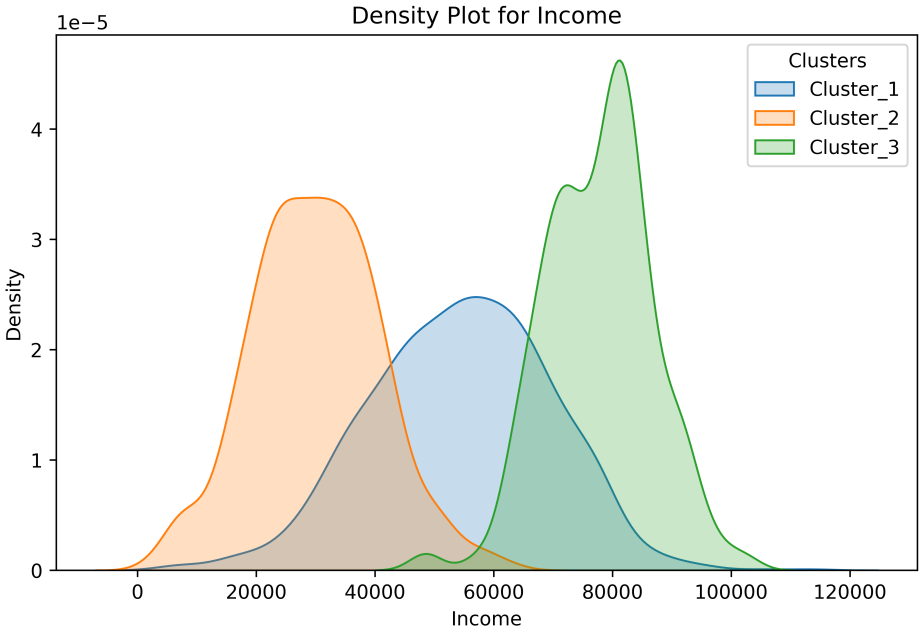
Segmentation

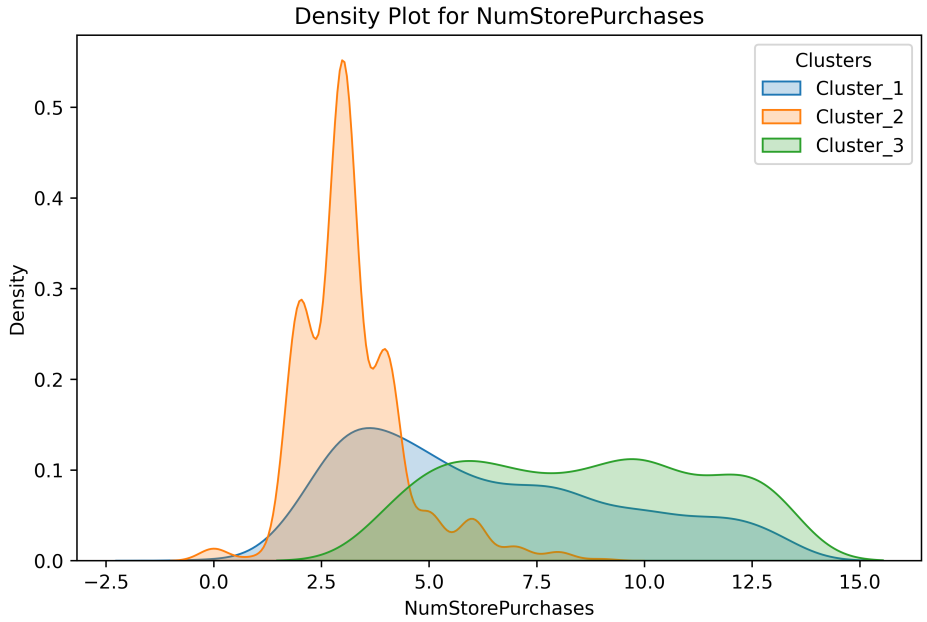
SOM application and analysis of segments

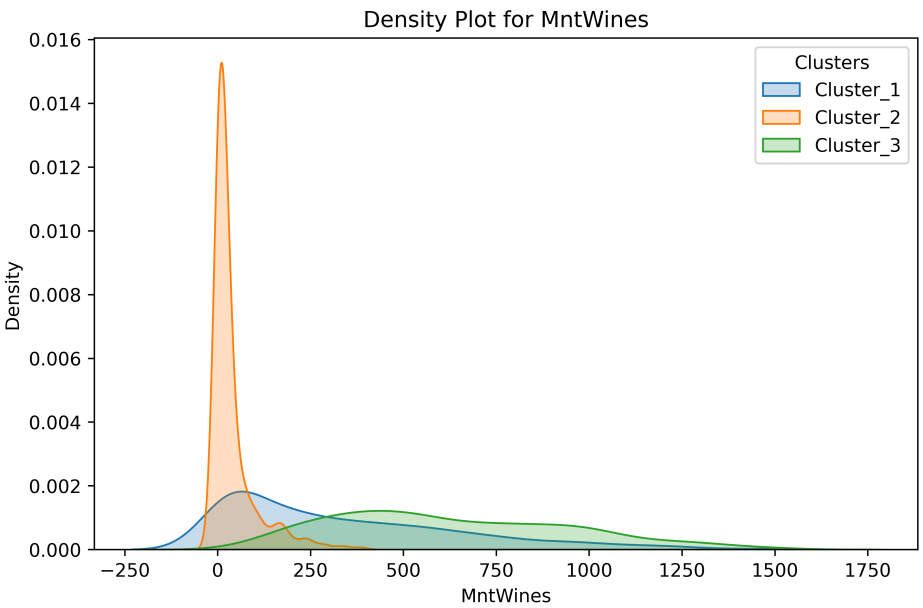
SOM results

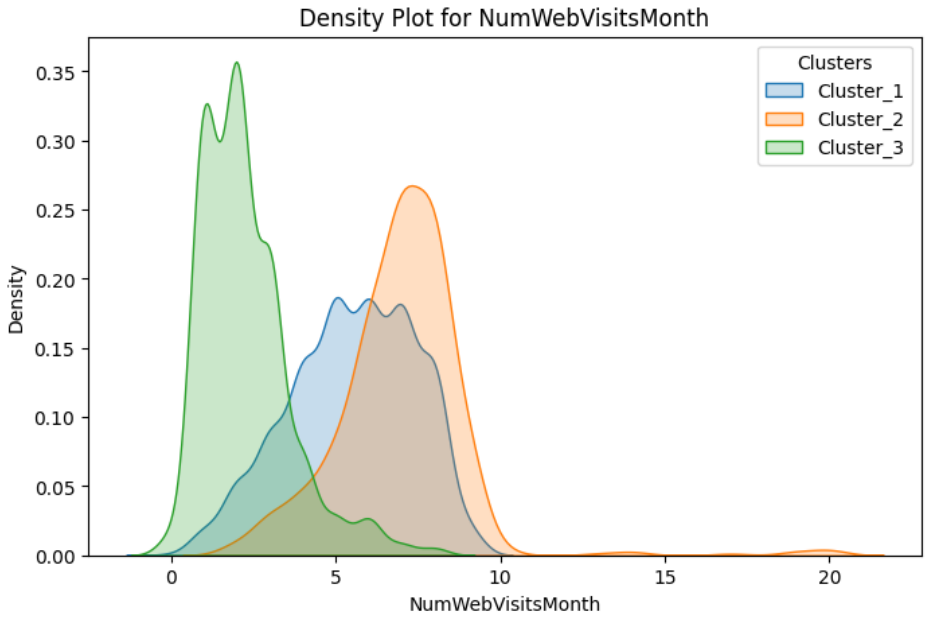

SOM Clusters
Clusters
Cluster 1 = 1189 obs
Cluster 2 = 630 obs
Cluster 3 = 393 obs
Cluster 4 = 4 obs
Income
Kidhome
Teenhome
MntWines
MntFruits
MntMeatProducts
MntFishProducts
NumWebPurchases
NumCatalogPurchases
NumStorePurchases
NumWebVisitsMonth

Customer Segmentation Analysis

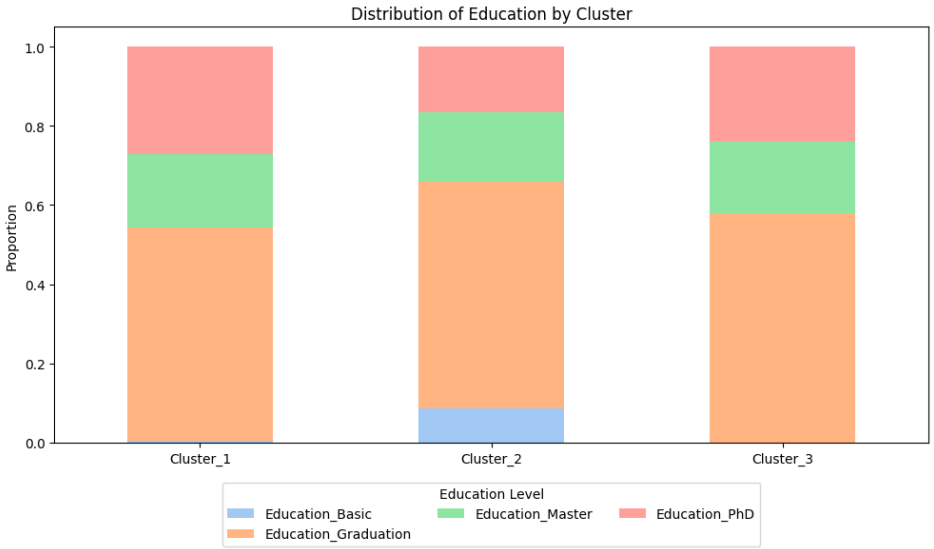


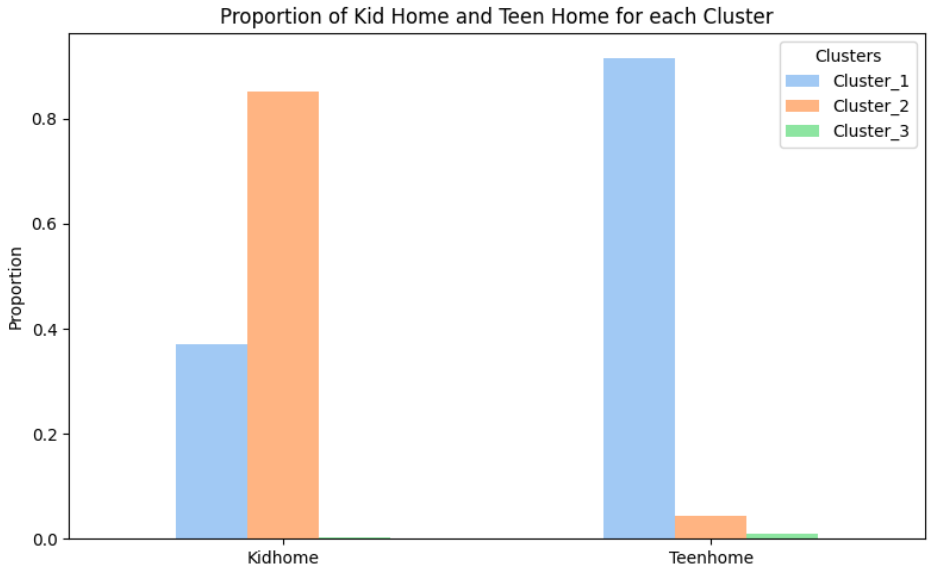




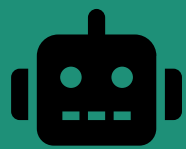


Customer Segmentation Analysis





- **Cluster 1**: Middle-income, balanced spending, moderate wine buyers.
- **Cluster 2**: Low-income, more children, low spending, frequent web visits.
- **Cluster 3**: High-income, few children, premium wine buyers.



Classification Models

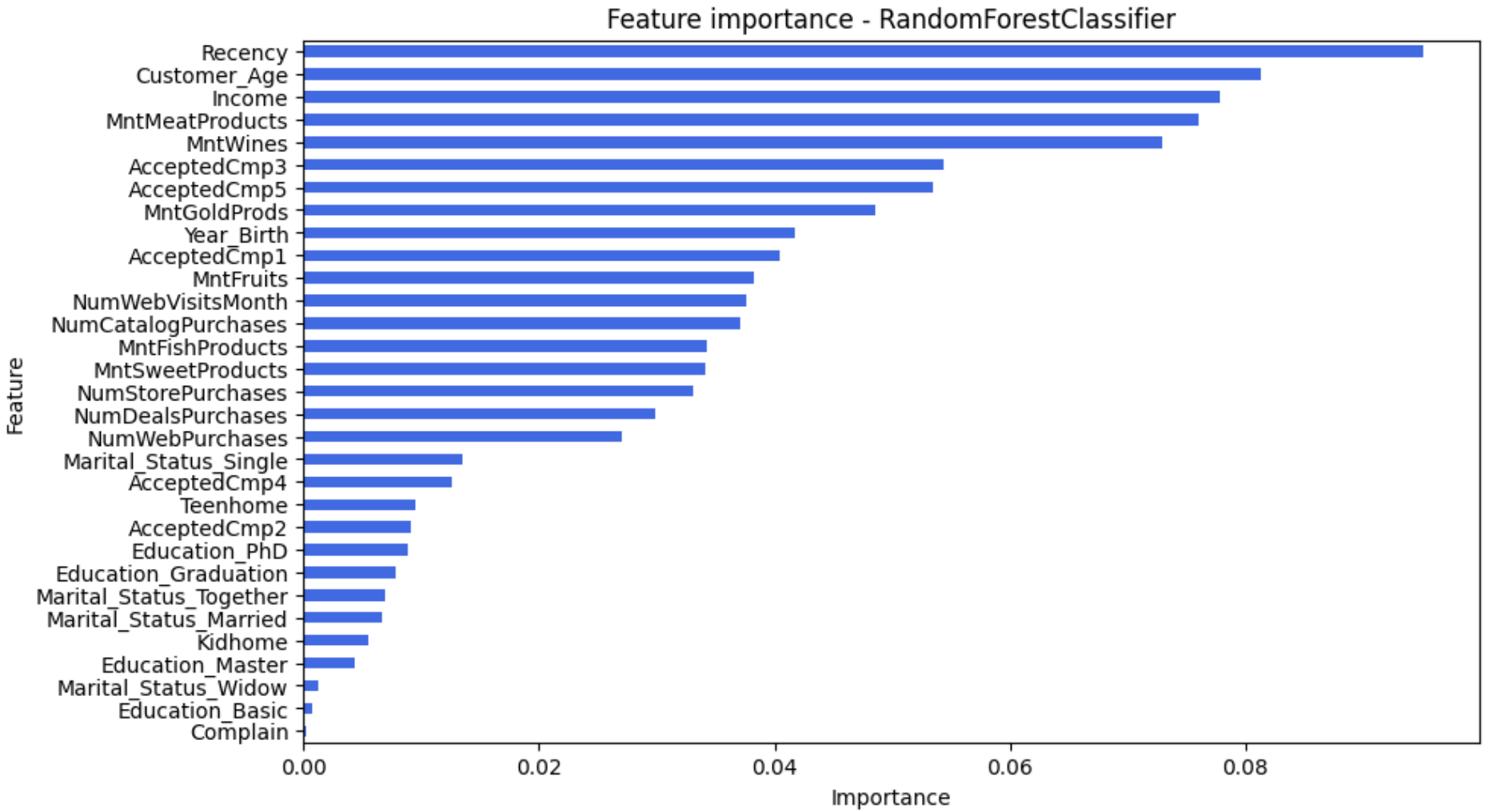
Gradient and logistic regression

Predictive Models

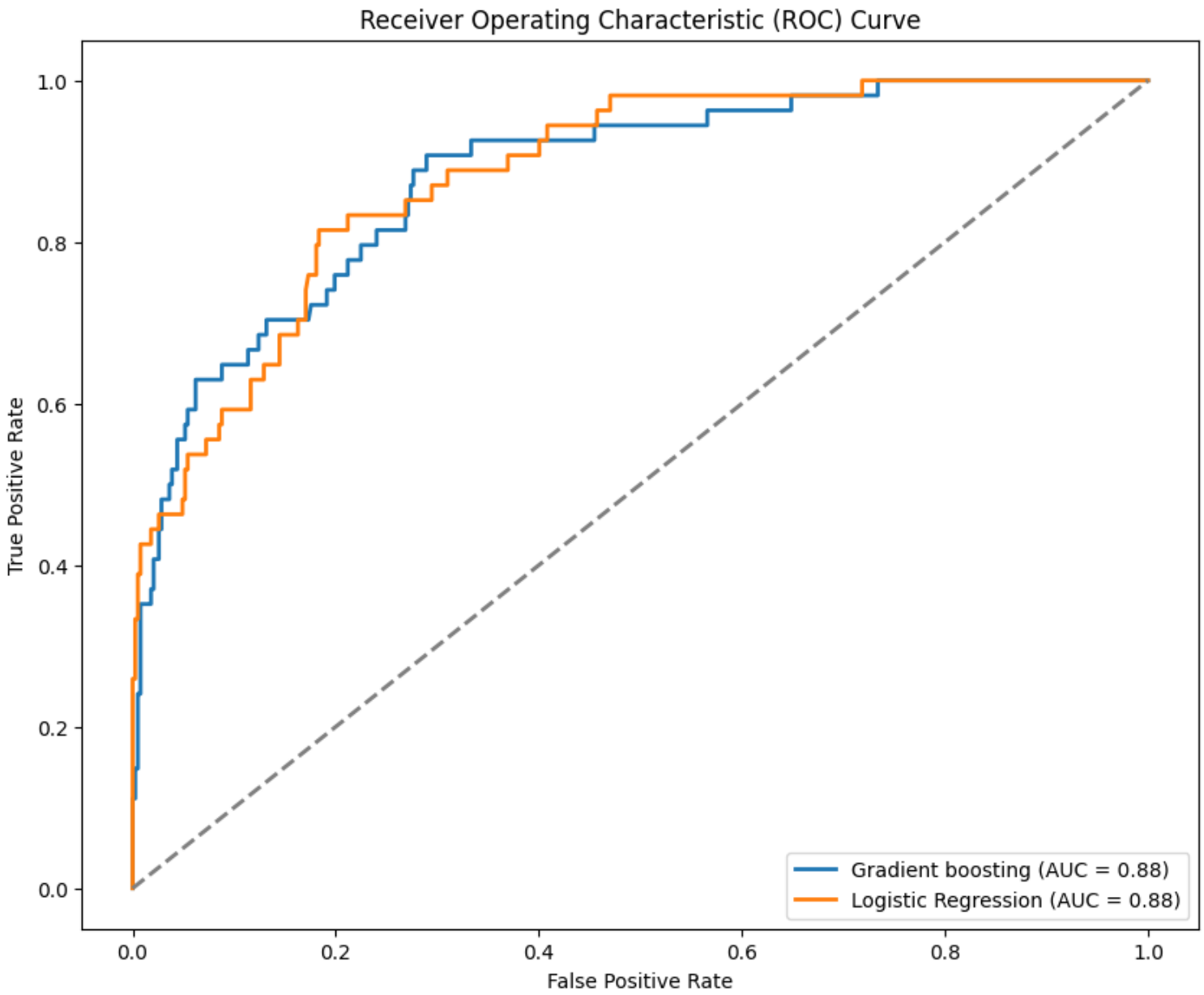
Test on two models (Logistic and Gradient Boosting):

- Accuracy: each $\approx 90\%$
- Key Factors:
  - Past campaign history
  - Customer income
  - Loyalty (Customer_Age and recency)
  - Consumption habits

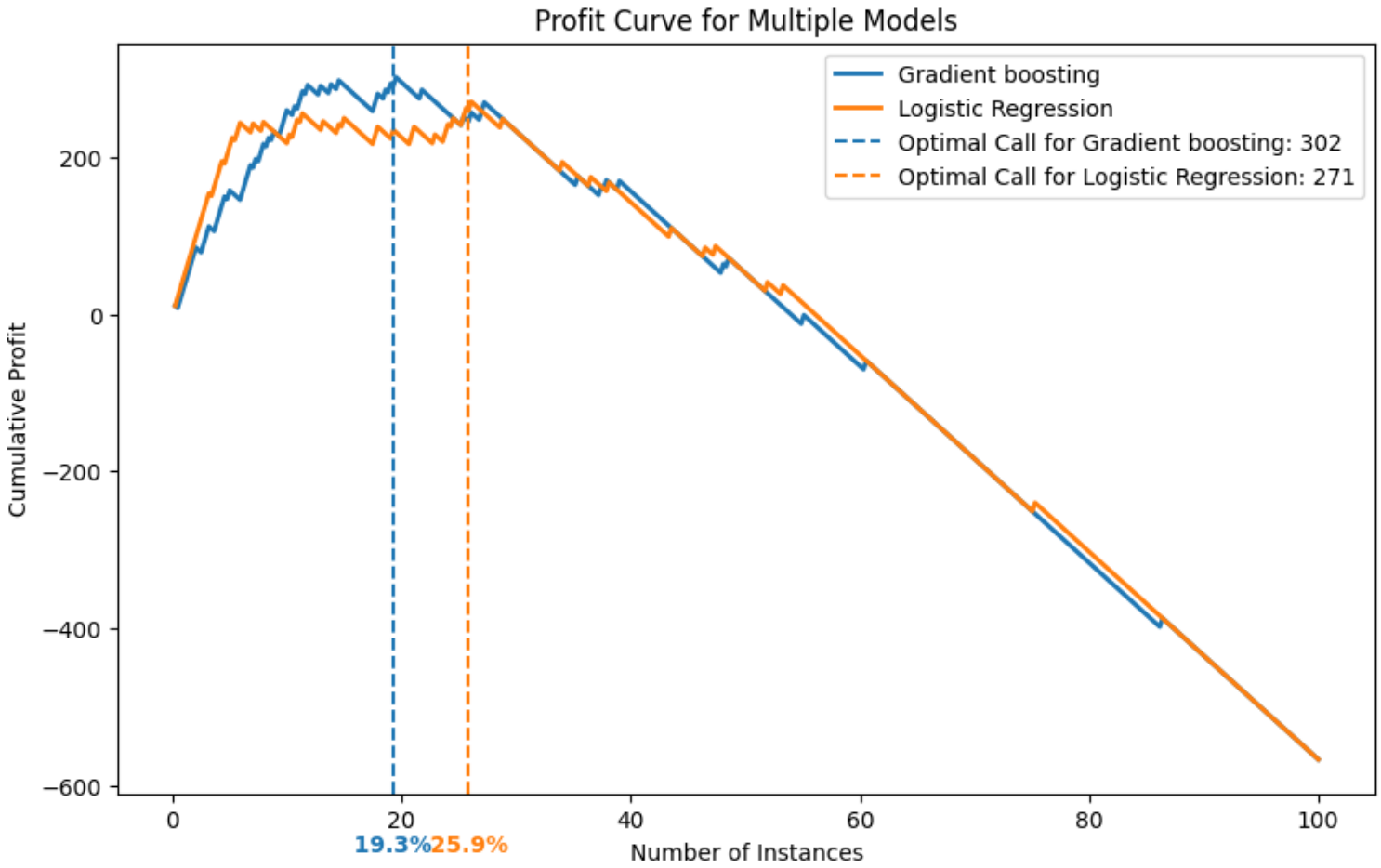
Feature Importance for tree



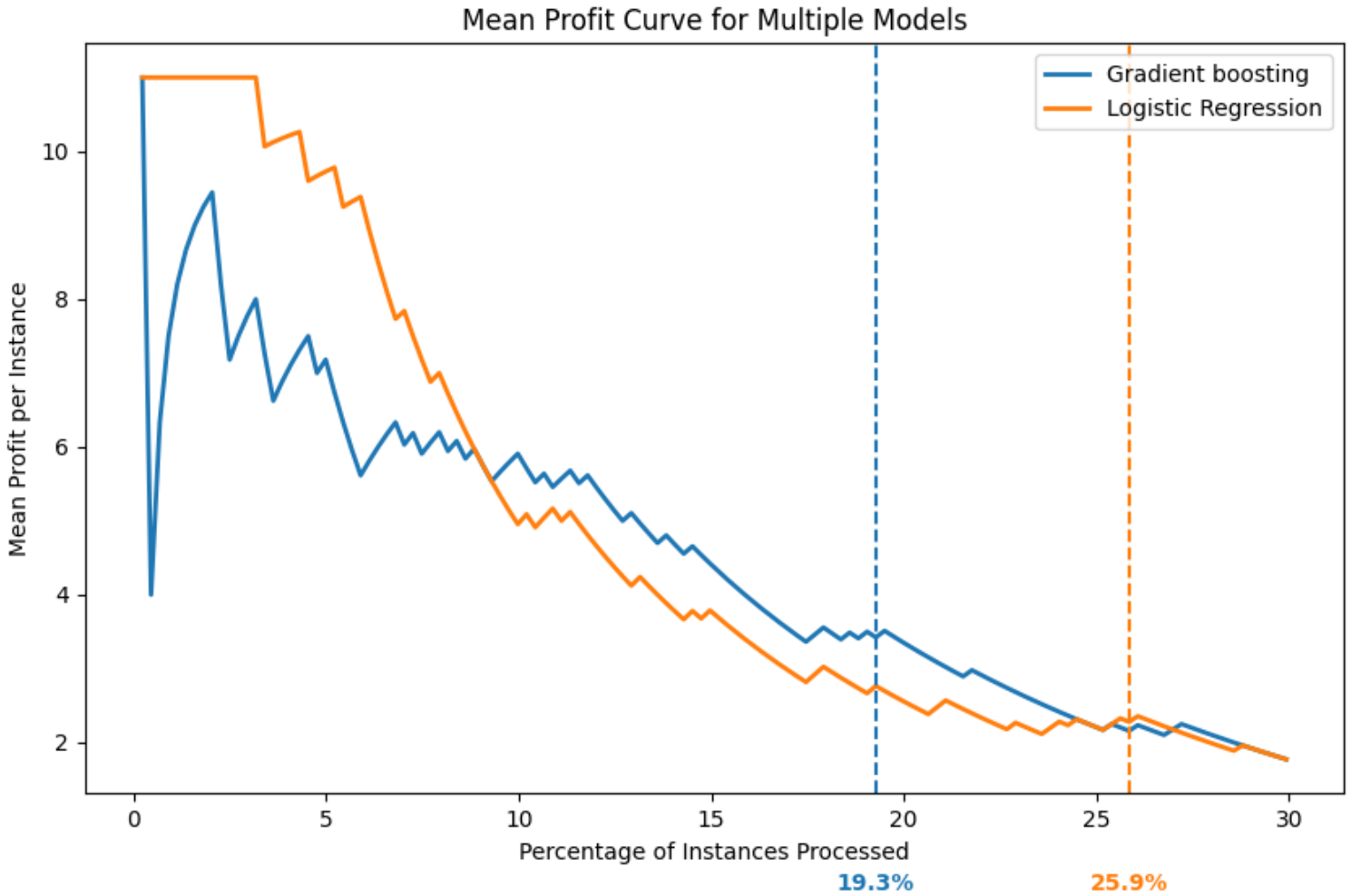
ROC and AUC


Receiver Operating Characteristic (ROC) Curve
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
Gradient boosting (AUC = 0.88)
Logistic Regression (AUC = 0.88)

€ Profit curve



€ Profit curve



Conclusion

- We will select the top 19.3% of consumers who are most likely to respond positively to the call, according to the gradient boosting model.
- At the optimal threshold of 19.3%, the average profit is 3.24MU compared to -1.36MU without data-driven decision-making.
- With our model, the total profit of the last campaign would have been 302.