

SUPERSTORE MARKETING CAMPAIGN : CUSTOMER DATA FOR ANALYSIS OF A TARGETED MEMBERSHIP OFFER

HENDRI KURNIAWAN



BUSINESS CONTEXT

Sebuah superstore ingin melaksanakan diskon akhir tahun. Untuk mendorong minat pembeli, mereka melakukan campaign dengan menawarkan gold membership yang memberikan diskon 20% untuk semua pembelian dengan harga 499 USD, dan akan naik menjadi 999 USD di lain hari khusus untuk existing customers lewat telepon. Pihak manajemen merasa cara terbaik untuk mengurangi biaya campaign adalah dengan cara membuat model yang dapat memprediksi customer yang diklasifikasikan akan membeli membership tersebut.



OBJECTIVES

- Memberikan prediksi terhadap variabel response yang membeli membership
- Merekomendasikan target customer yang seperti apa yang harus di hubungi untuk campaign selanjutnya



INPUT VARIABELS

- 1.ID - Unique ID of each customer
 - 2.Year_Birth - Age of the customer
 - 3.Education - customer's level of education
 - 4.Marital - customer's marital status
 - 5.Income - customer's yearly household income
 - 6.Kidhome - number of small children in customer's household
 - 7.Teenhome - number of teenagers in customer's household
 - 8.Dt_Customer - date of customer's enrollment with the company
 - 9.Recency - number of days since the last purchase
 - 10.MntWines - the amount spent on wine products in the last 2 years
 - 11.MntFruits - the amount spent on fruits products in the last 2 years
 - 12.MntMeatProducts - the amount spent on meat products in the last 2 years
 - 13.MntFishProducts - the amount spent on fish products in the last 2 years
 - 14.MntSweetProducts - amount spent on sweet products in the last 2 years
 - 15.MntGoldProds - the amount spent on gold products in the last 2 years
 - 16.NumDealsPurchases - number of purchases made with discount
 - 17.NumWebPurchases - number of purchases made through the company's website
 - 18.NumCatalogPurchases - number of purchases made using catalog (buying goods to be shipped through the mail)
 - 19.NumStorePurchases - number of purchases made directly in stores
 - 20.NumWebVisitsMonth - number of visits to company's website in the last month
 - 21.Complain - 1 if the customer complained in the last 2 years,0 otherwise
- Target Variable :**
- 1.Response (target) - 1 if customer accepted the offer in the last campaign, 0 otherwise



DATASET INFORMATION

2240
BARIS

22
KOLOM

1 TARGET
RESPONSE

24 BARIS
MISSING
VALUE
INCOME

0
DUPLICATED
VALUES

BUSINESS QUESTIONS

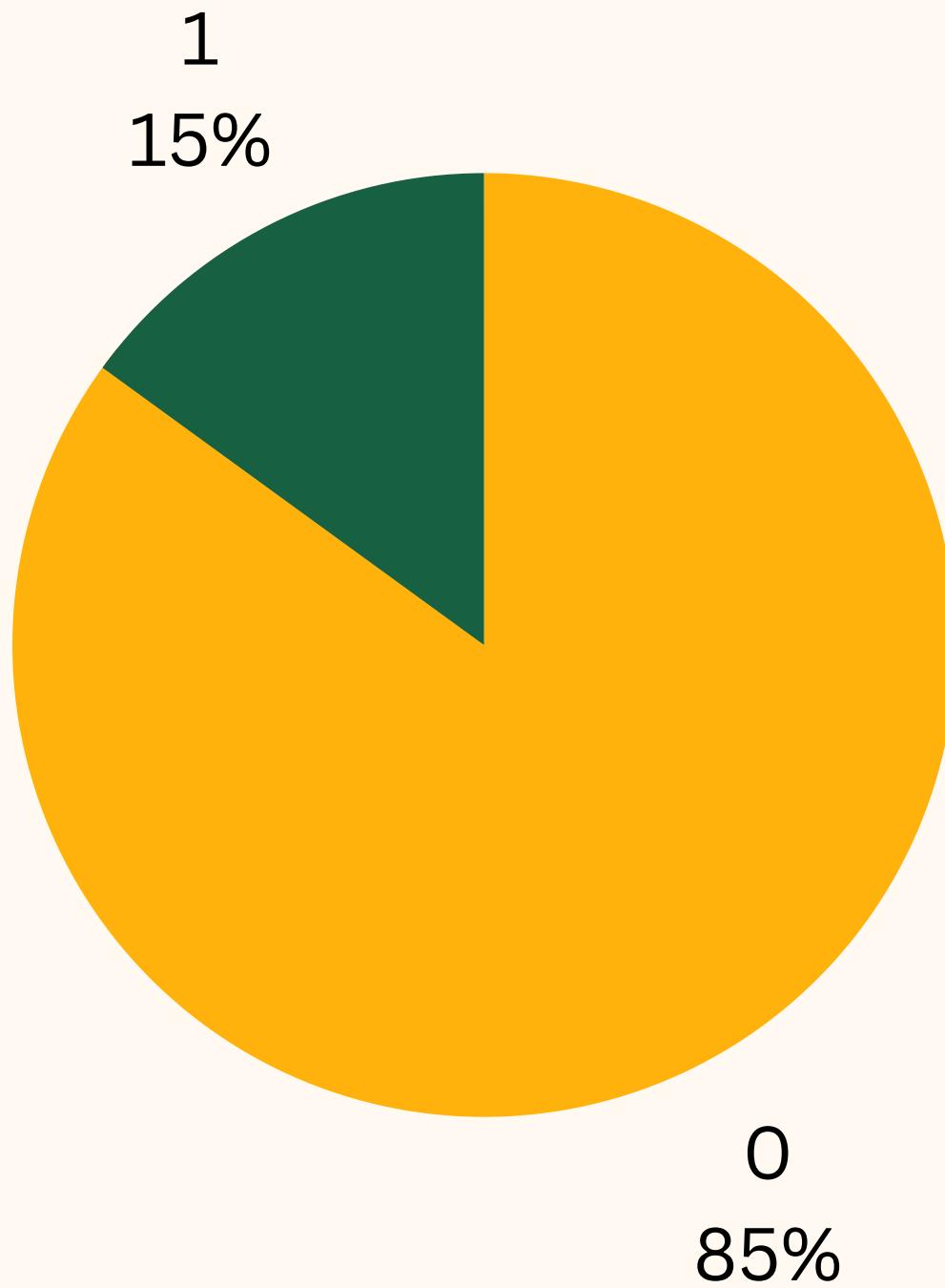
1. Bagaimana Insight dari variabel Profil Customer terhadap Response?
2. Bagaimana pengaruh jenis produk terhadap Response?
3. Berapa banyak pembelian yang dilakukan lewat marketing channel yang tersedia di superstore, dan bagaimana pengaruhnya terhadap variable response?
4. Apakah semakin banyak uang yang dihabiskan oleh customer mempengaruhi pembelian membership?
5. Apakah semakin tinggi income mempengaruhi total pengeluaran untuk belanja?
6. Bagaimana tren nominal pembelanjaan konsumen di Superstore dari tahun 2012 - 2014? dan Bagaimana tren mingguannya?



EXPLORATORY DATA ANALYSIS

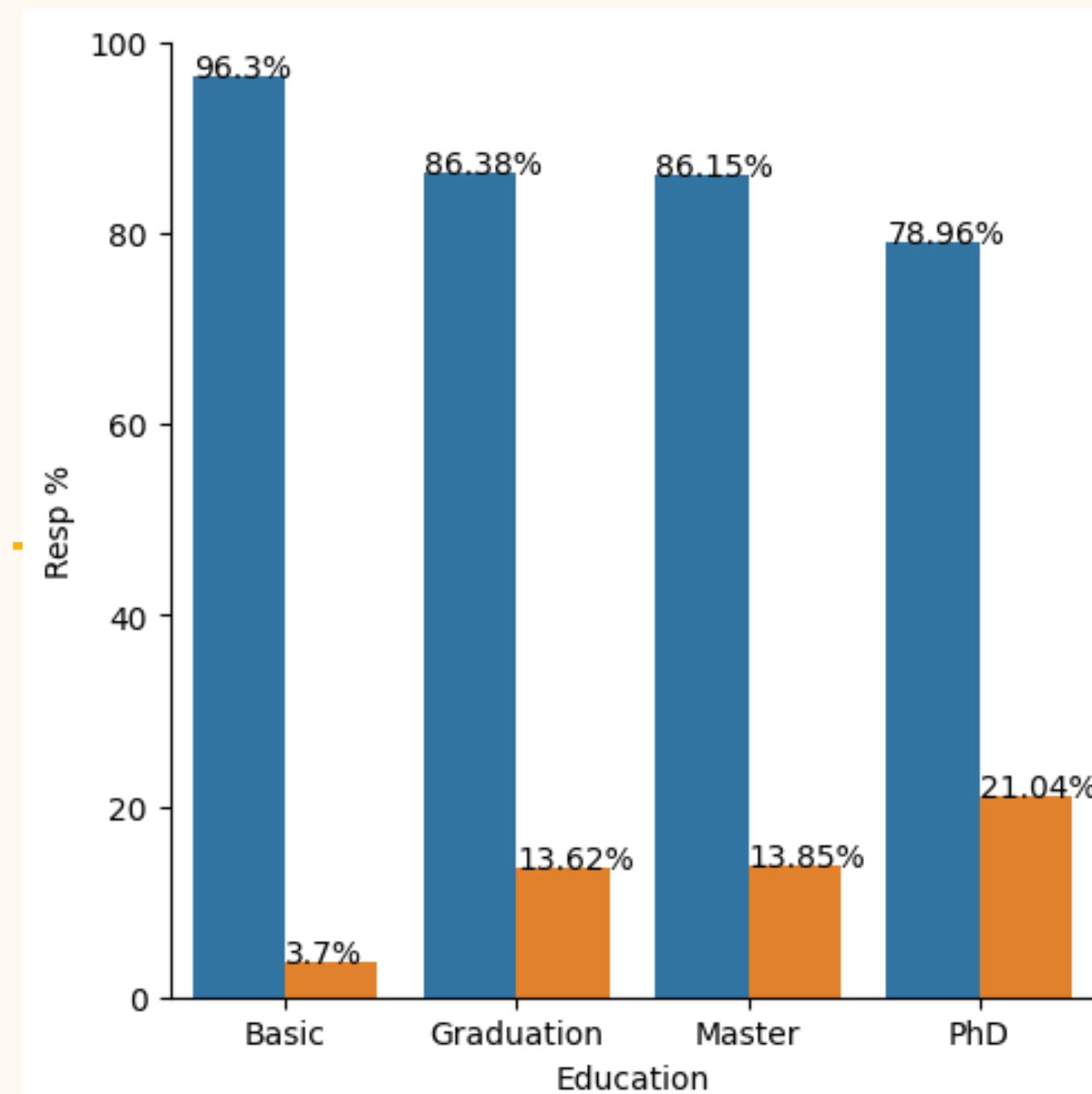


TARGET VARIABEL

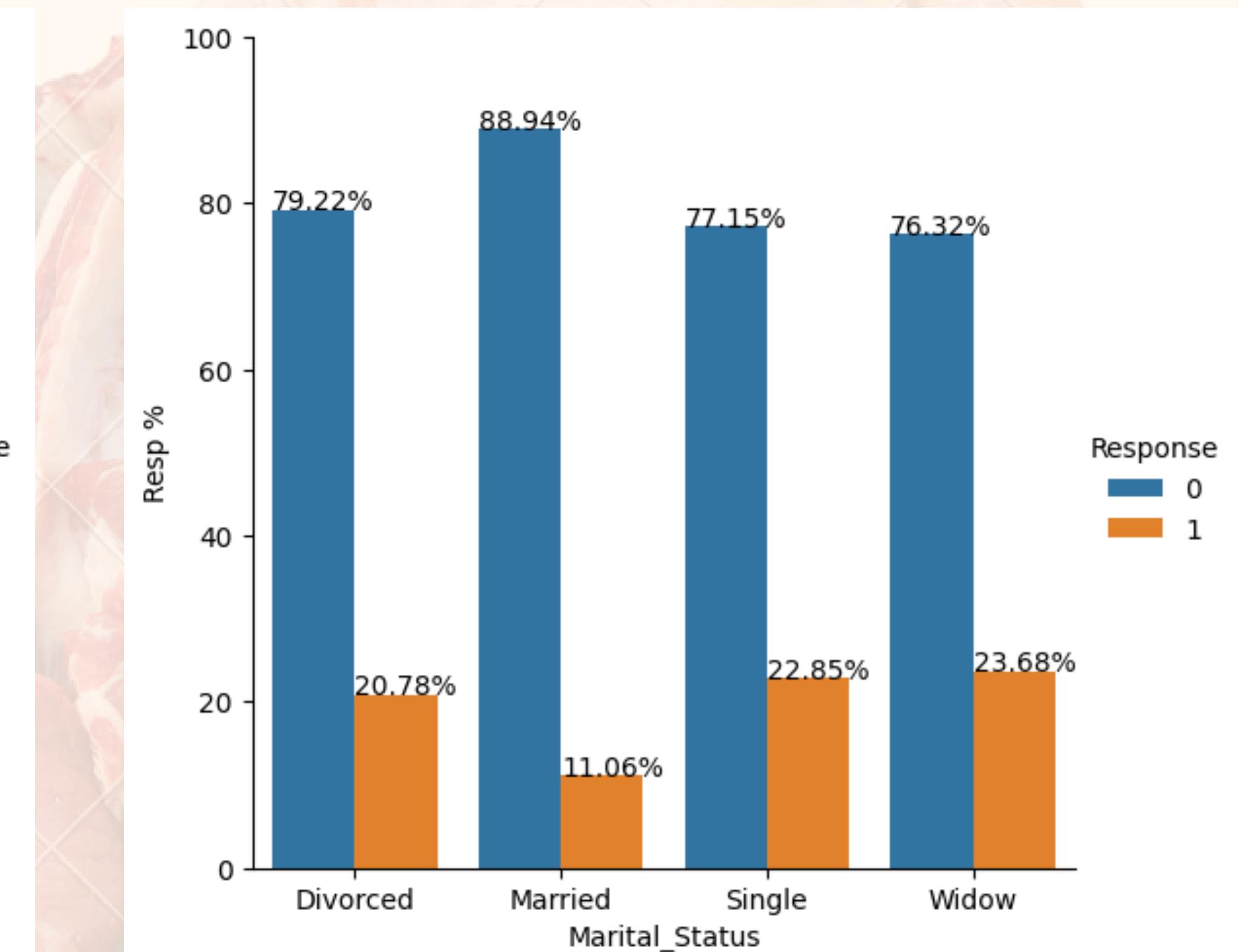


Persentasi pembelian membership hanya sebanyak 15%, sementara 85% tidak memiliki membership

EDUCATION AND MARITAL STATUS VS RESPONSE

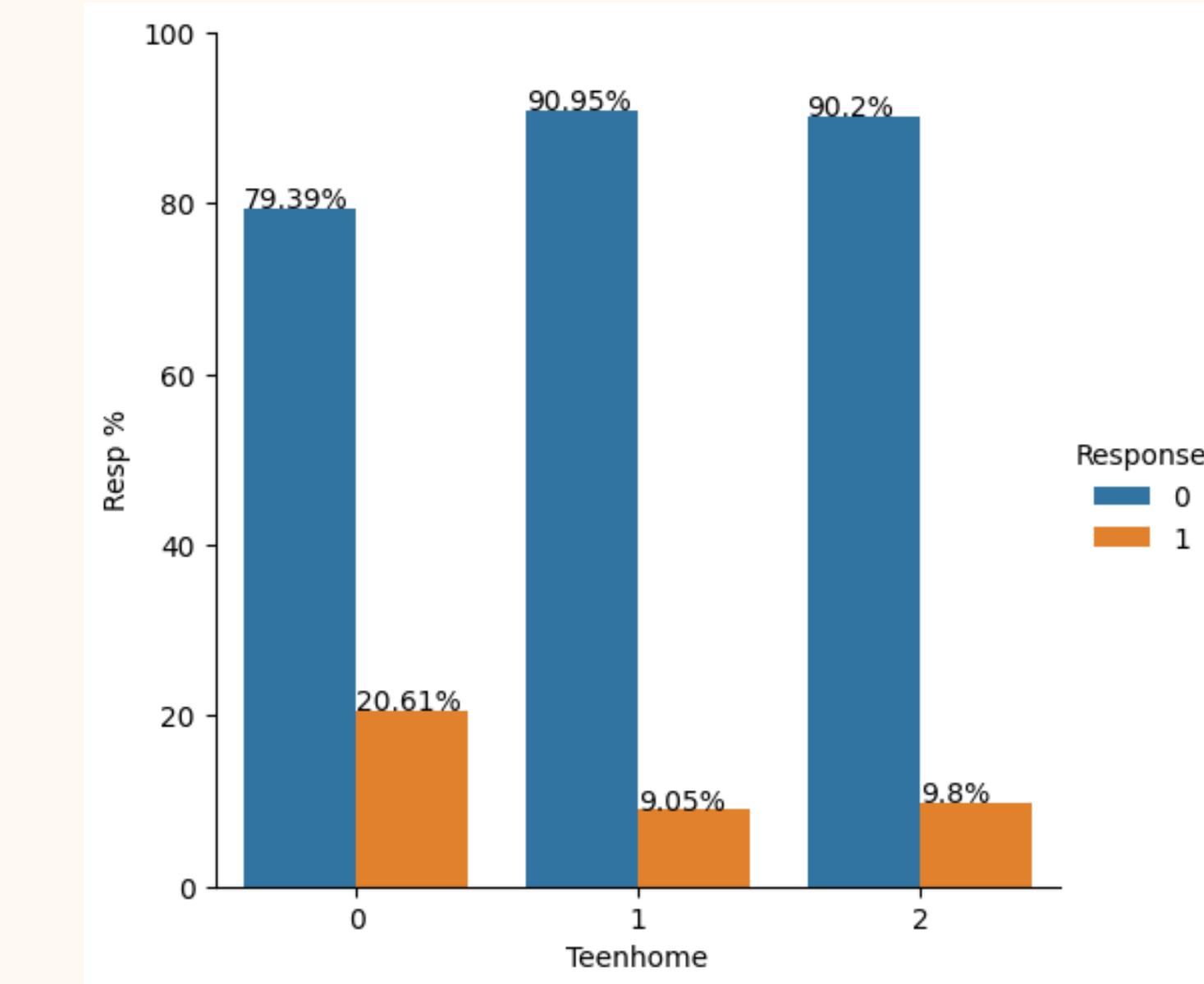
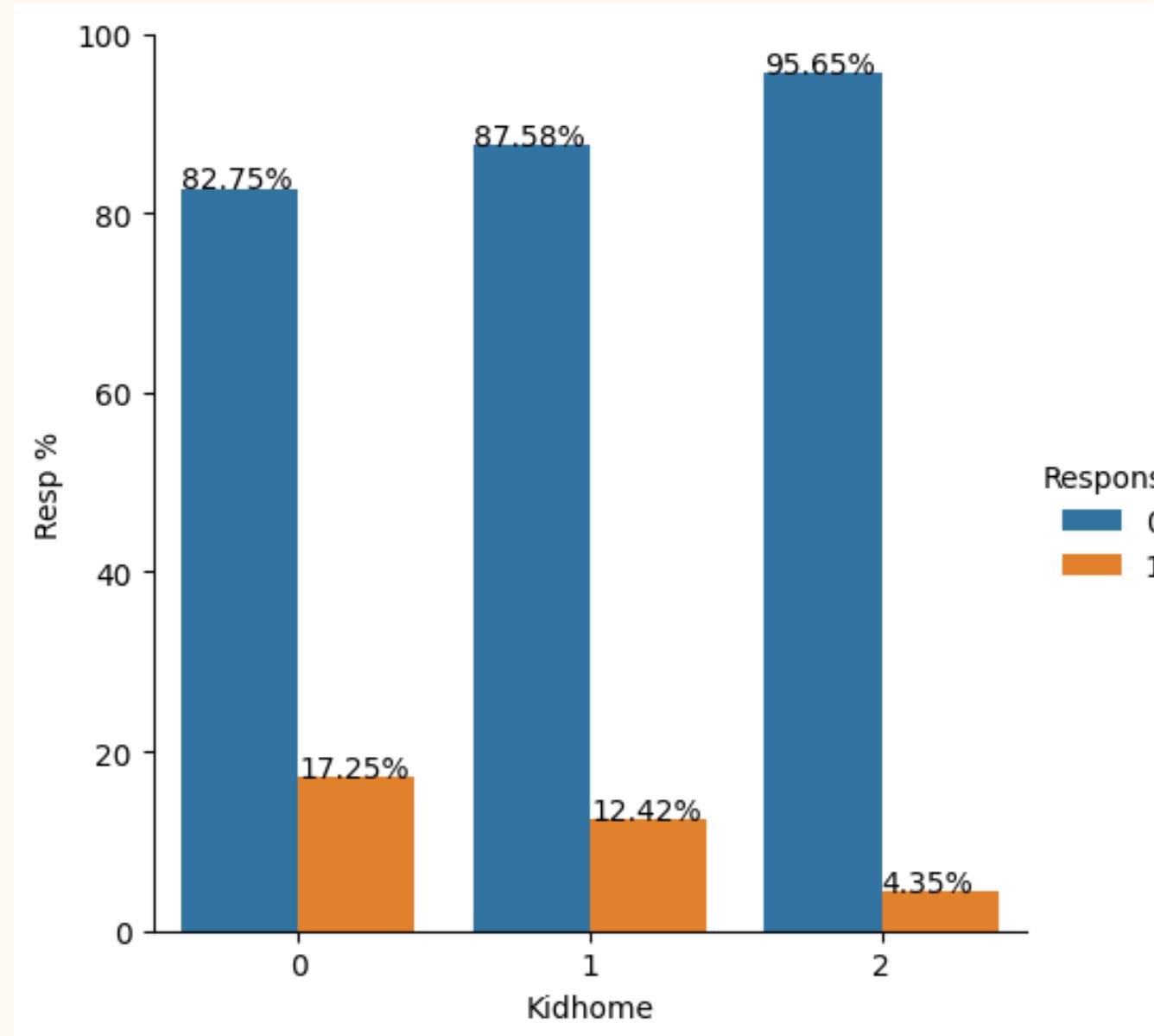


Semakin tinggi Education customer, semakin banyak pula yang membeli membership di superstore. Dan sebaliknya, semakin rendah tingkat pendidikan, semakin sedikit yang membeli membership.



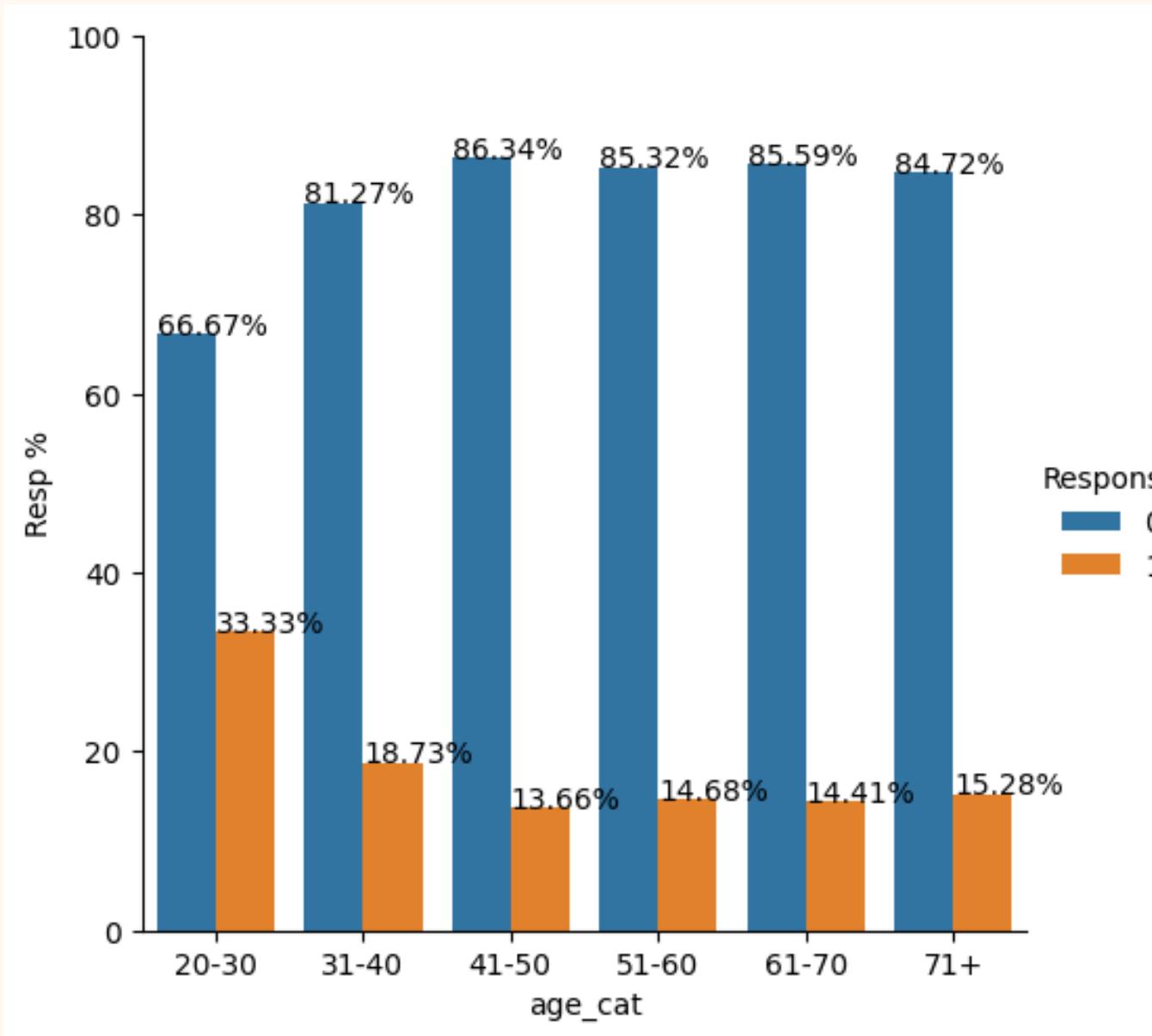
Customer dengan Marital Status (Widow) merupakan customer yang paling banyak melakukan pembelian membership dengan persentase sekitar 23.68 %

KIDHOME AND TEENHOME VS RESPONSE

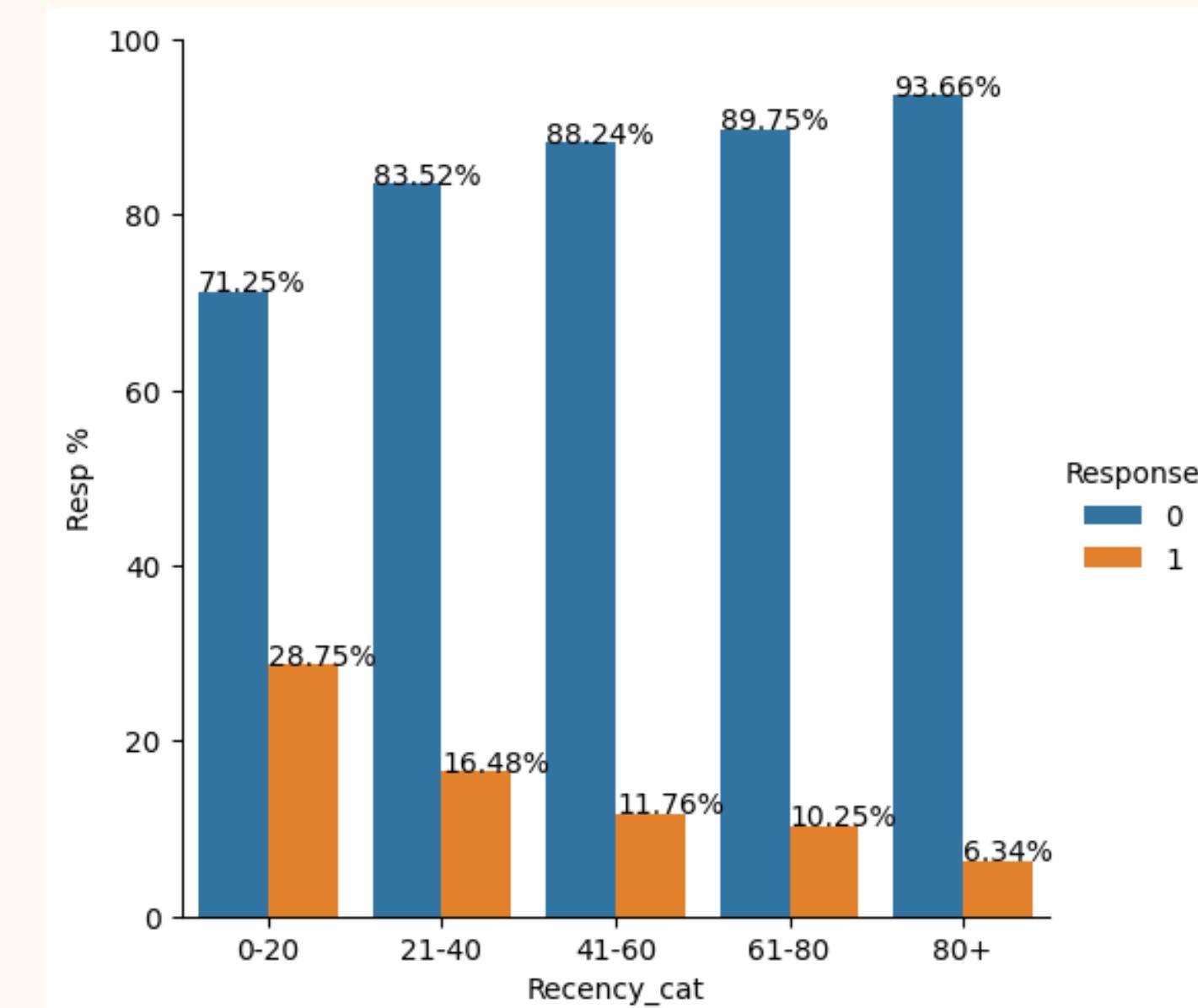


Customer yang sama sekali tidak memiliki Kidhome dan Teenhome cenderung membeli membership

AGE AND RECENCY VS RESPONSE

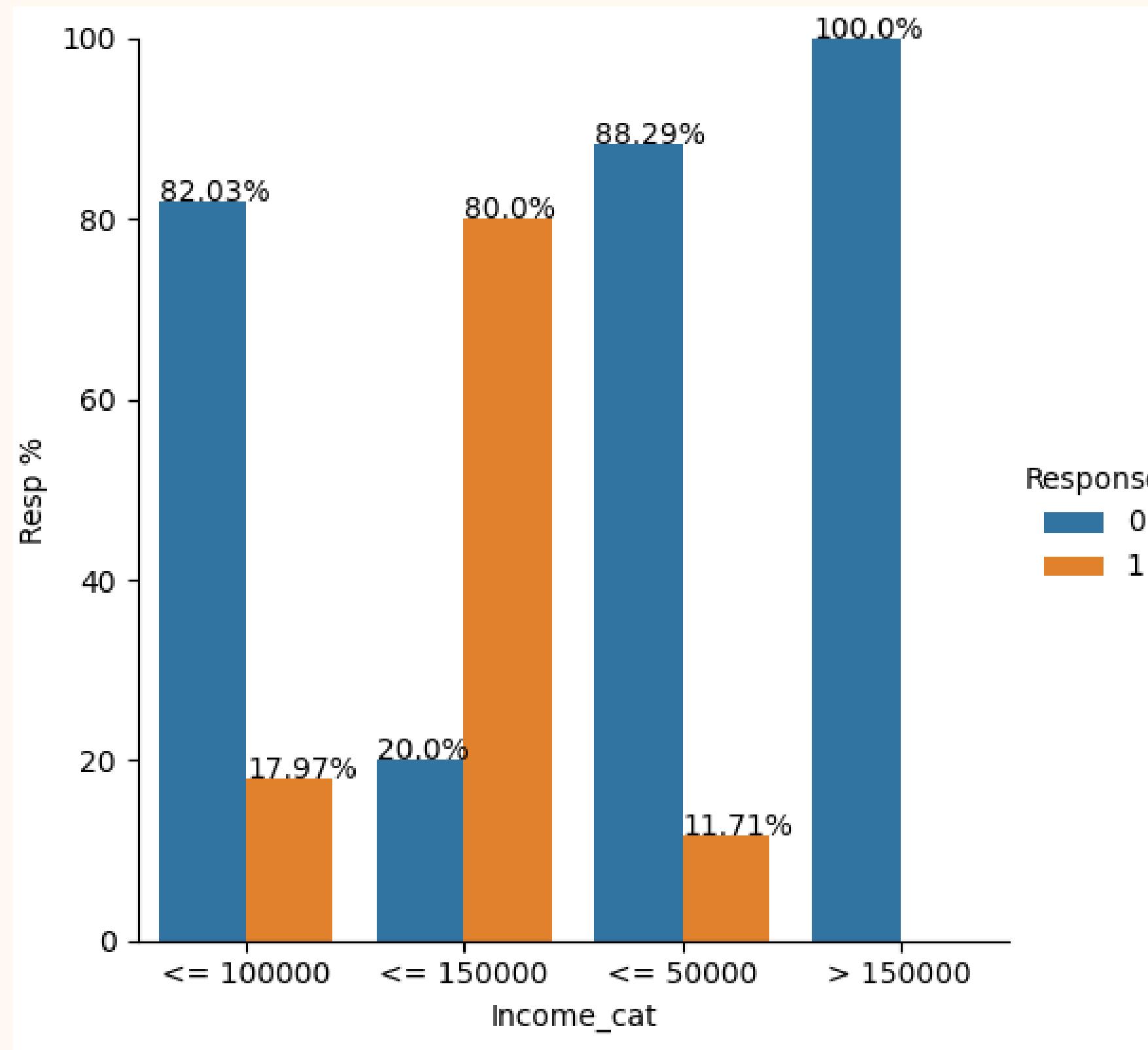


Jumlah customer dengan pembeli membership terbanyak ada di jarak umur antara 20 - 30 tahun (33.33%)



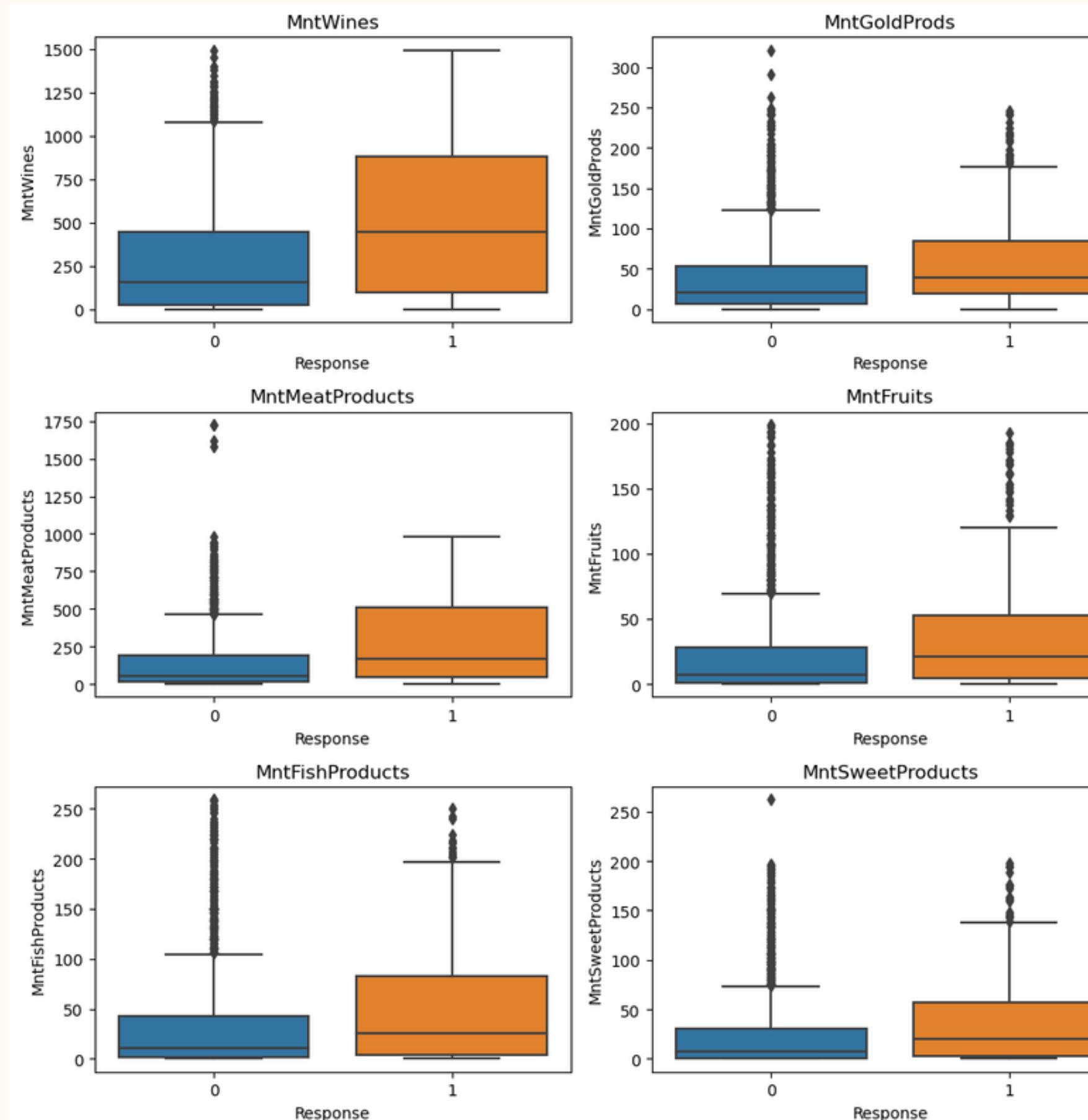
Berdasarkan insight diatas, customer dengan jangka waktu pembelian di range 0 - 20 hari dari tanggal pembelian sebelumnya melakukan pembelian membership paling banyak (28.75%). Artinya semakin dekat jarak waktu suatu pembelian dengan tanggal pembelian terakhir, customer cenderung membeli membership dan sebaliknya.

INCOME VS RESPONSE



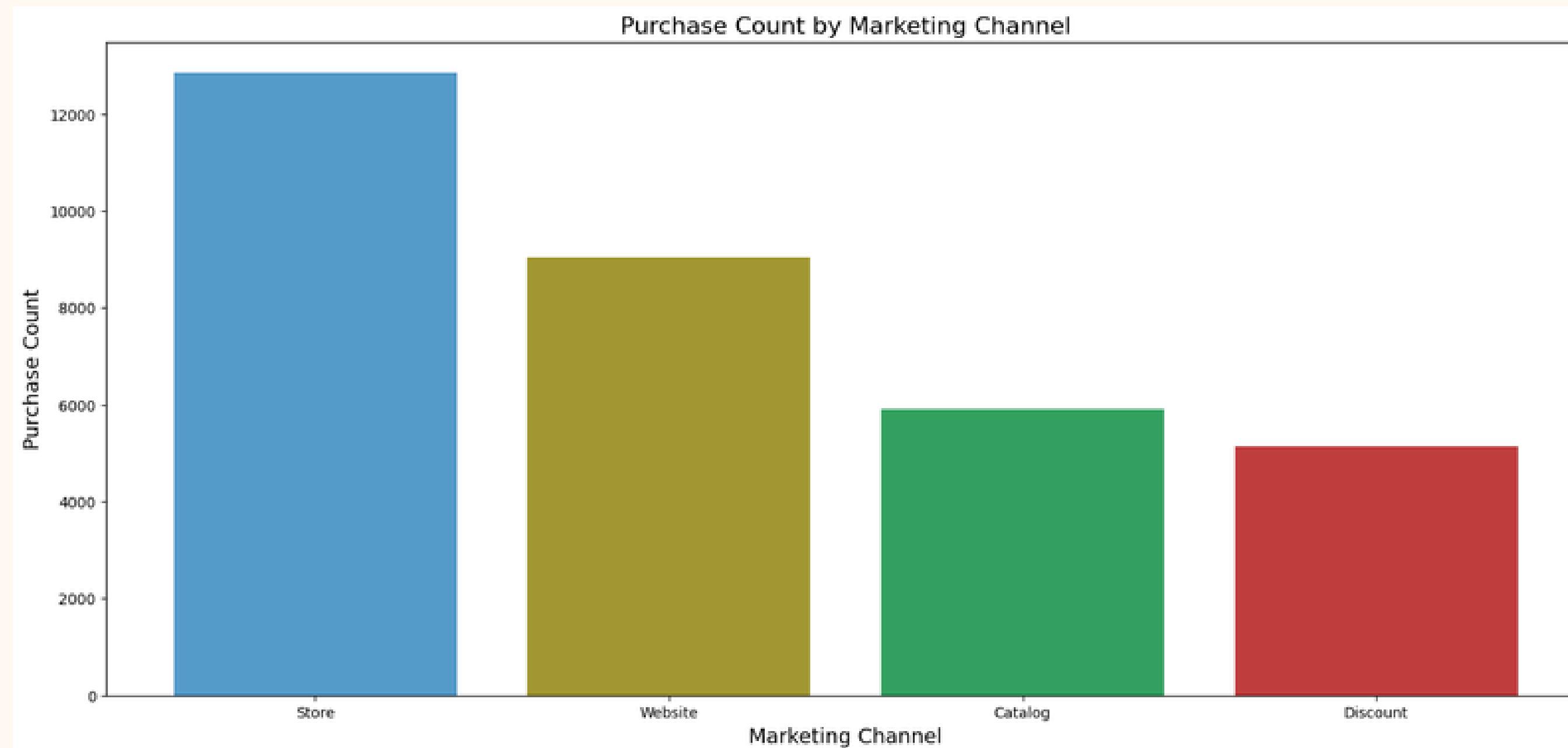
Secara persentase jumlah customer dengan kategori Income 100.000 - 150.000 melakukan pembelian membership sebanyak 80% . Dengan demikian, istilah semakin banyak income customer semakin juga mendorong penjualan membership di superstore ini hanya berlaku pada range income di angka 0 - 150.000. karena terdapat customer dengan income > 150.000 namun dia tidak membeli membership sama sekali.

JENIS PRODUK VS RESPONSE



Masing - masing boxplot diatas menunjukan bahwa customer yang menghabiskan lebih banyak uangnya pada produk-produk diatas cenderung akan mengambil/membeli membership. Artinya produk-produk superstore yang ada memiliki peranan yang sama dan berpengaruh pada penjualan membership.

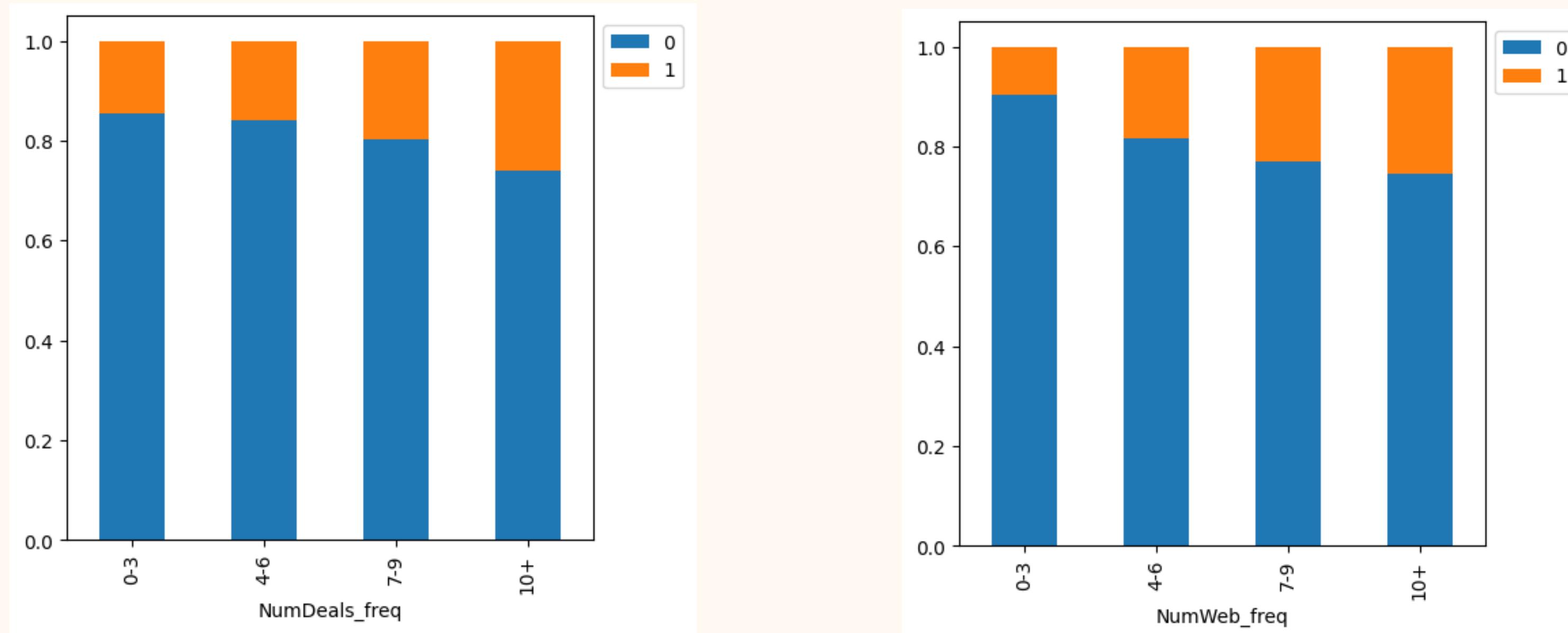
JUMLAH PEMBELIAN YANG DILAKUKAN DI MARKETING CHANNEL



Marketing Channel yang menghasilkan pembelian terbanyak adalah Store dengan jumlah 12.847 pembelian, diikuti oleh Website,Catalog & Discount dengan masing-masing sebanyak 9046, 5912 & 5146 pembelian.

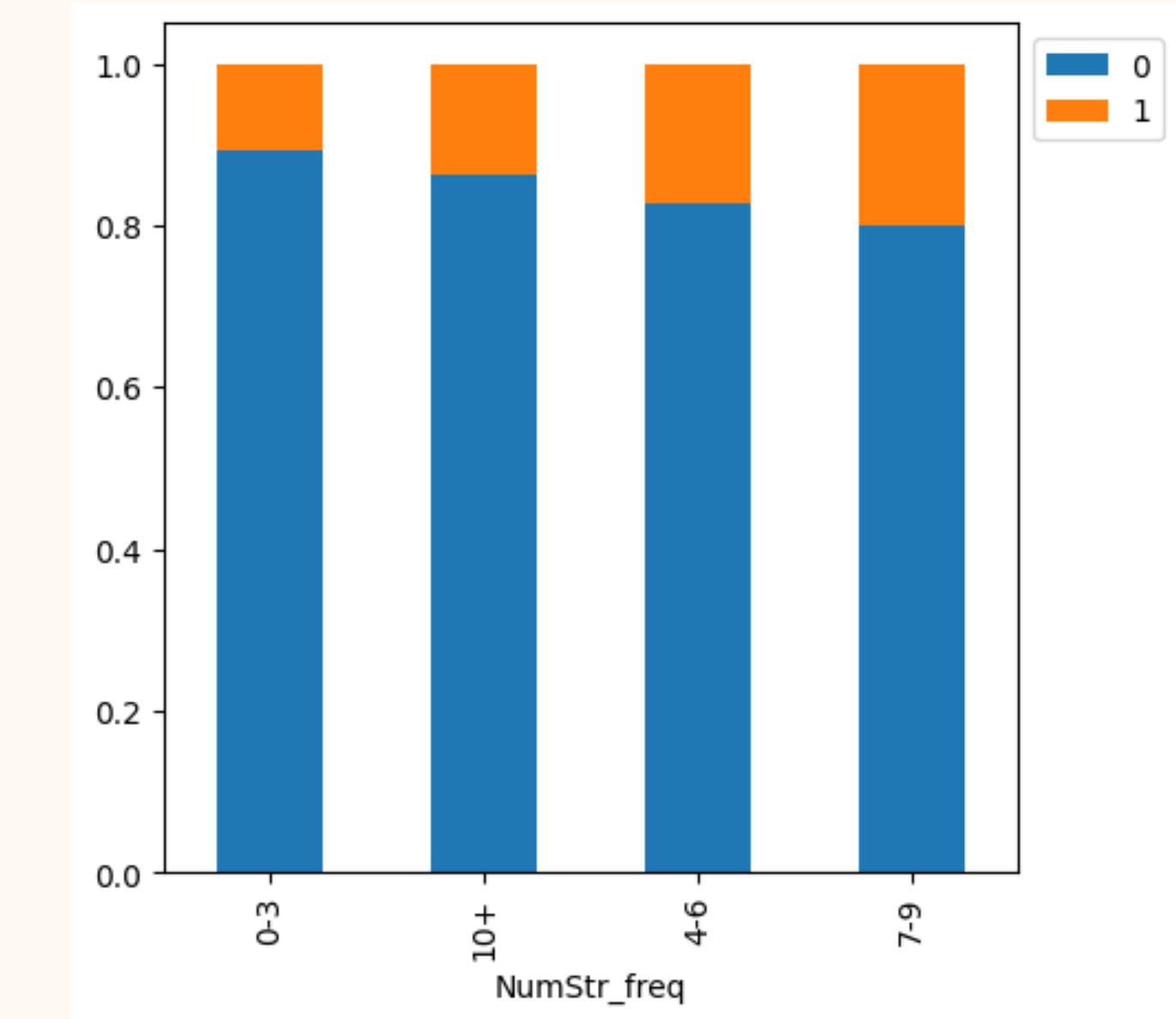
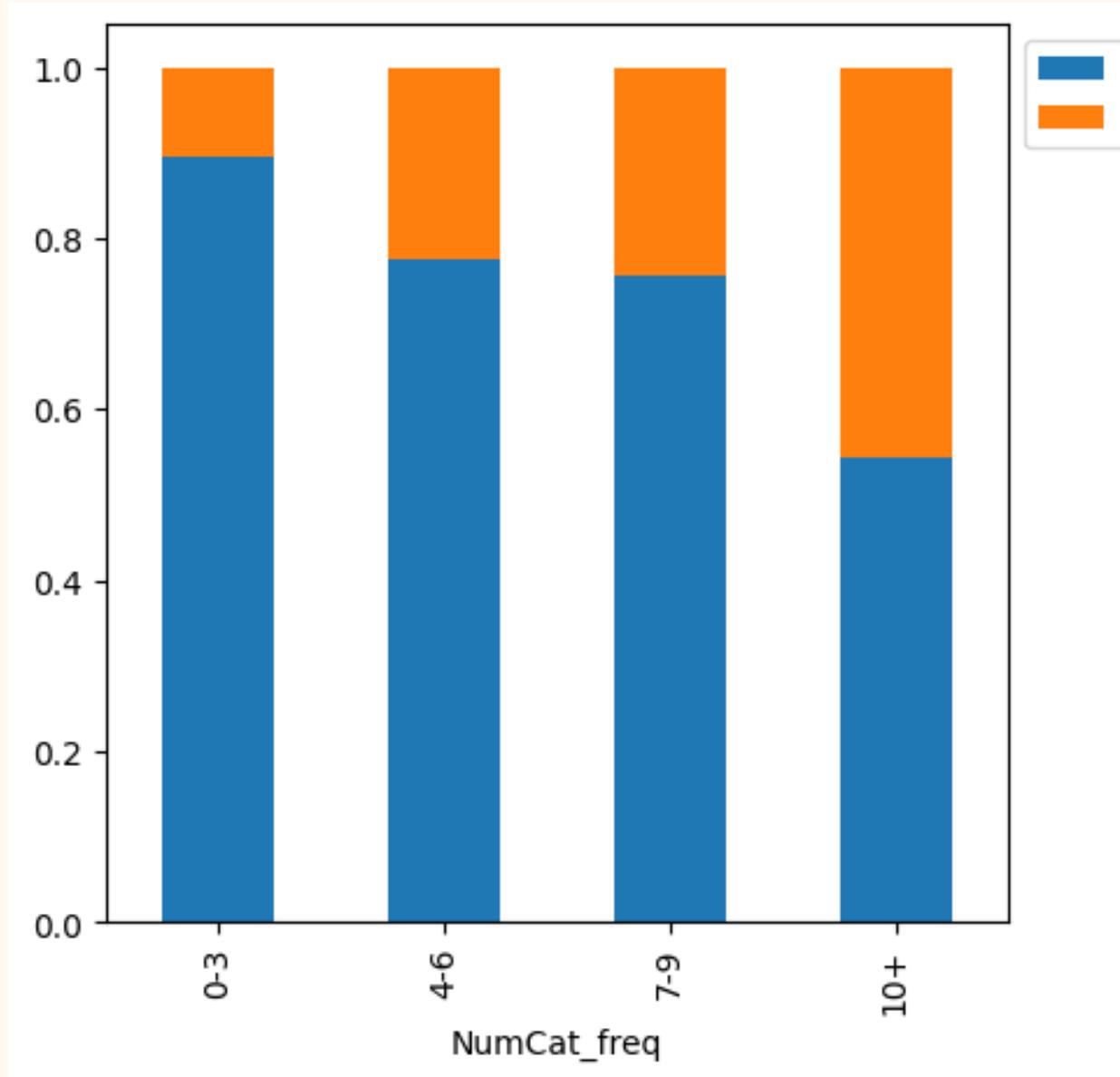


NUMDEALS PURCHASES AND NUMWEB PURCHASES VS RESPONSES



Pembelian di atas 10 x cenderung membuat customer membeli membership. Artinya Semakin banyak pembelian yang dilakukan oleh customer lewat Web (NumWebPurchases) dan diskon(NumDealsPurchases) ,semakin banyak pula customer yang membeli membership

NUMCATALOG PURCHASES AND NUMSTORE PURCHASES VS RESPONSES

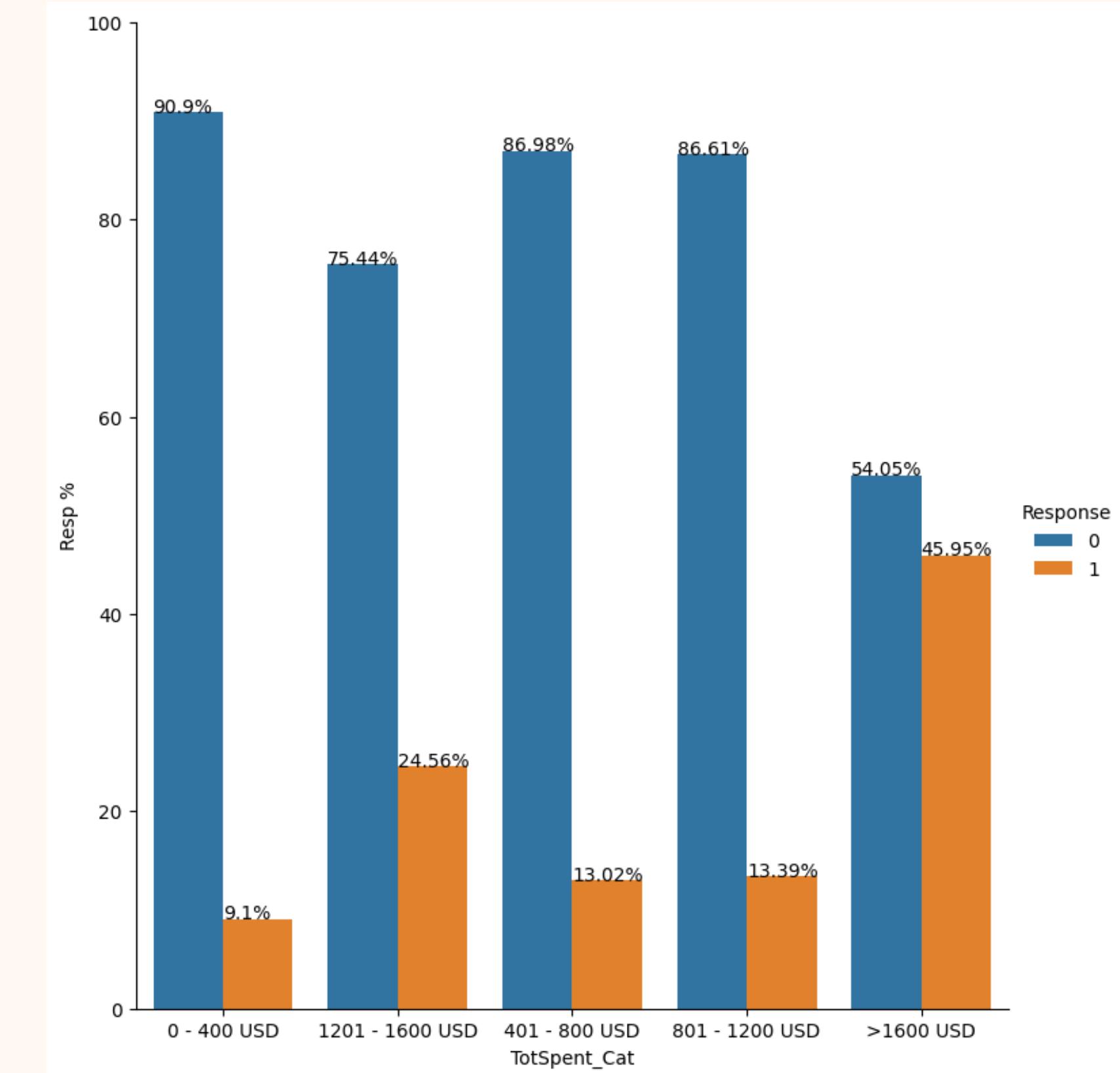


Masih selaras dengan data diatas sebelumnya, semakin banyak pembelian yang dilakukan, semakin banyak pula customer yang membeli membership.

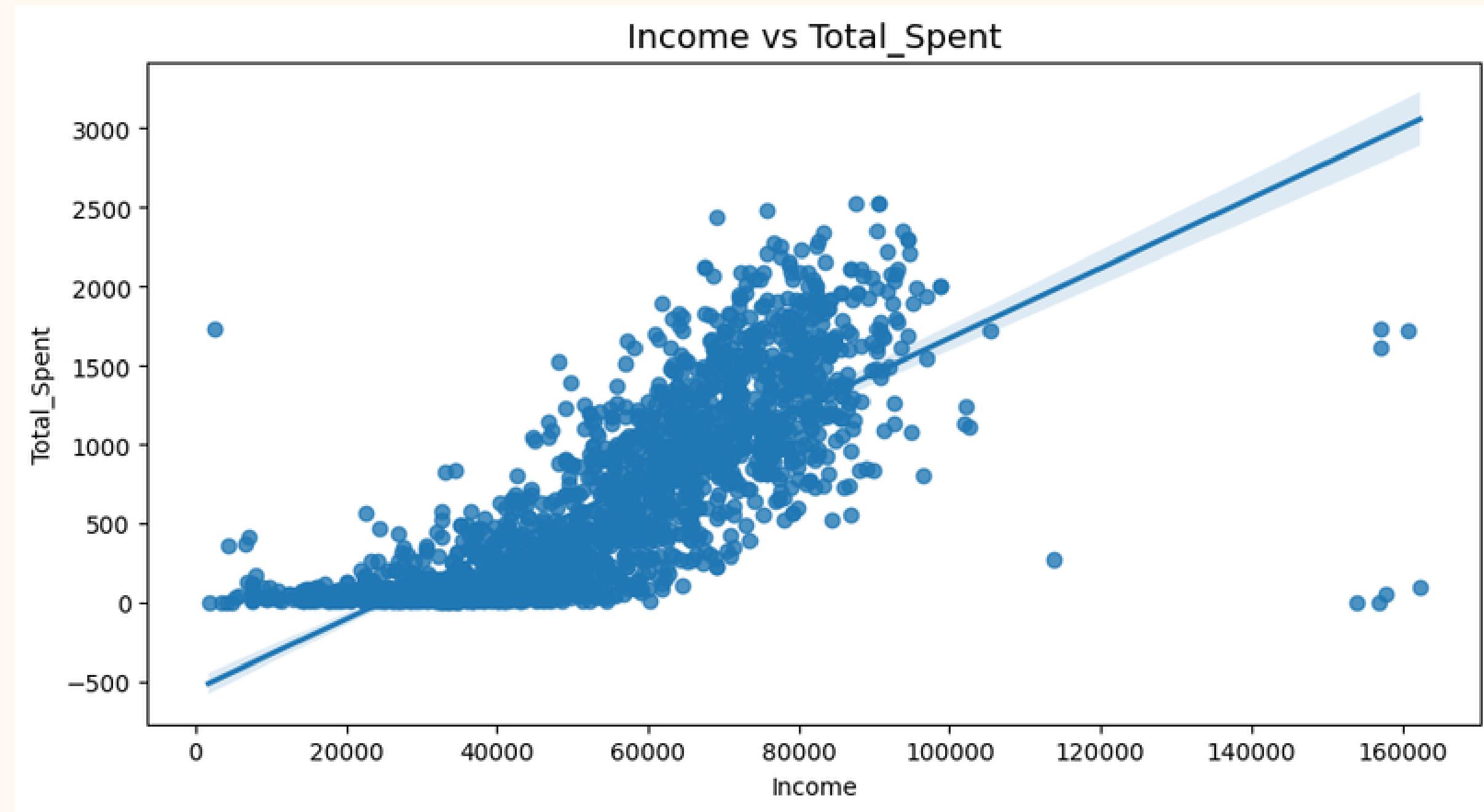


TOTAL PEMBELANJAAN VS RESPONSE

- Secara persentase customer dengan nominal pembelanjaan diatas 1600 USD memiliki tingkat rate membership yang sangat tinggi,hampir 50% atau setengahnya dari jumlah customer yang ada.
- Seiring nominal pembelanjaan membesar,persentase pembelian membership juga meningkat.

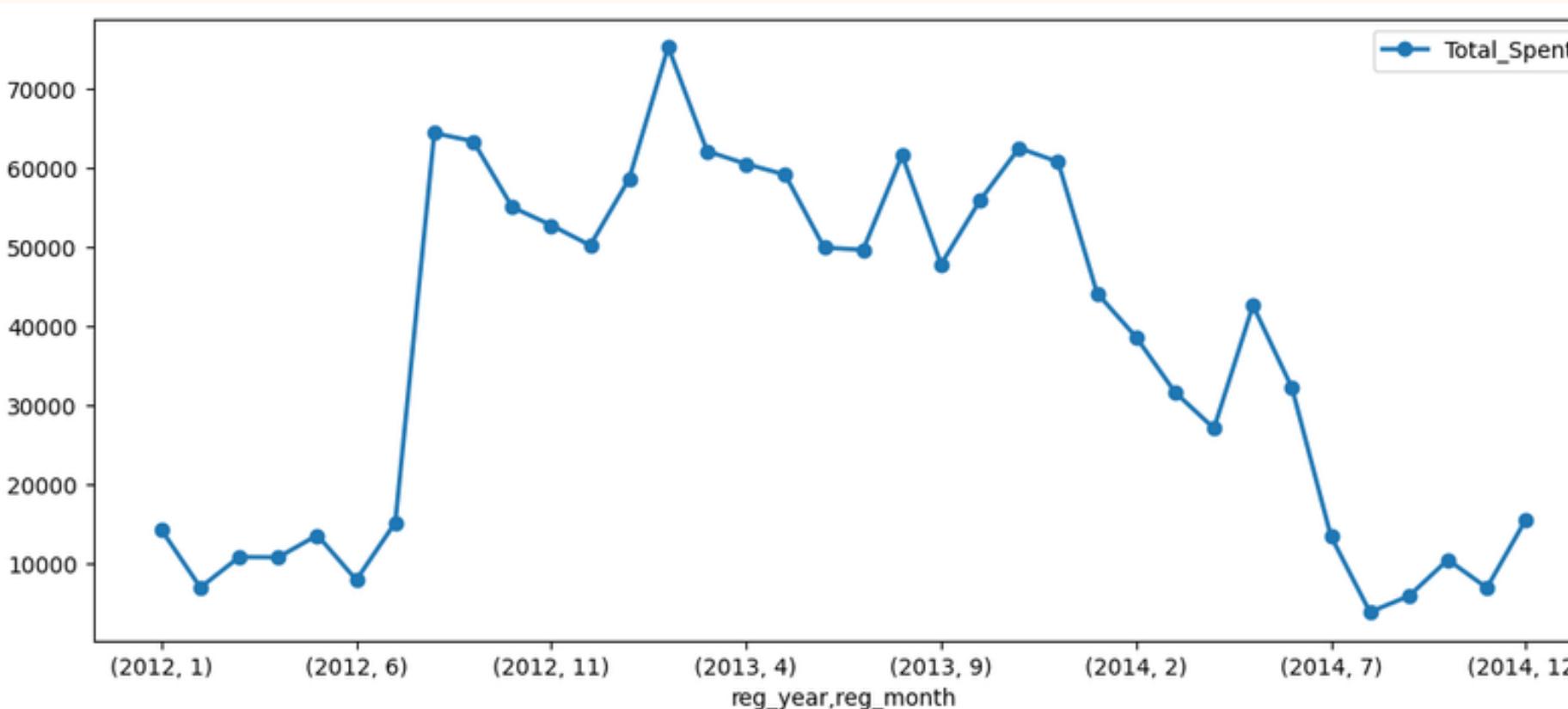


TOTAL INCOME VS RESPONSE

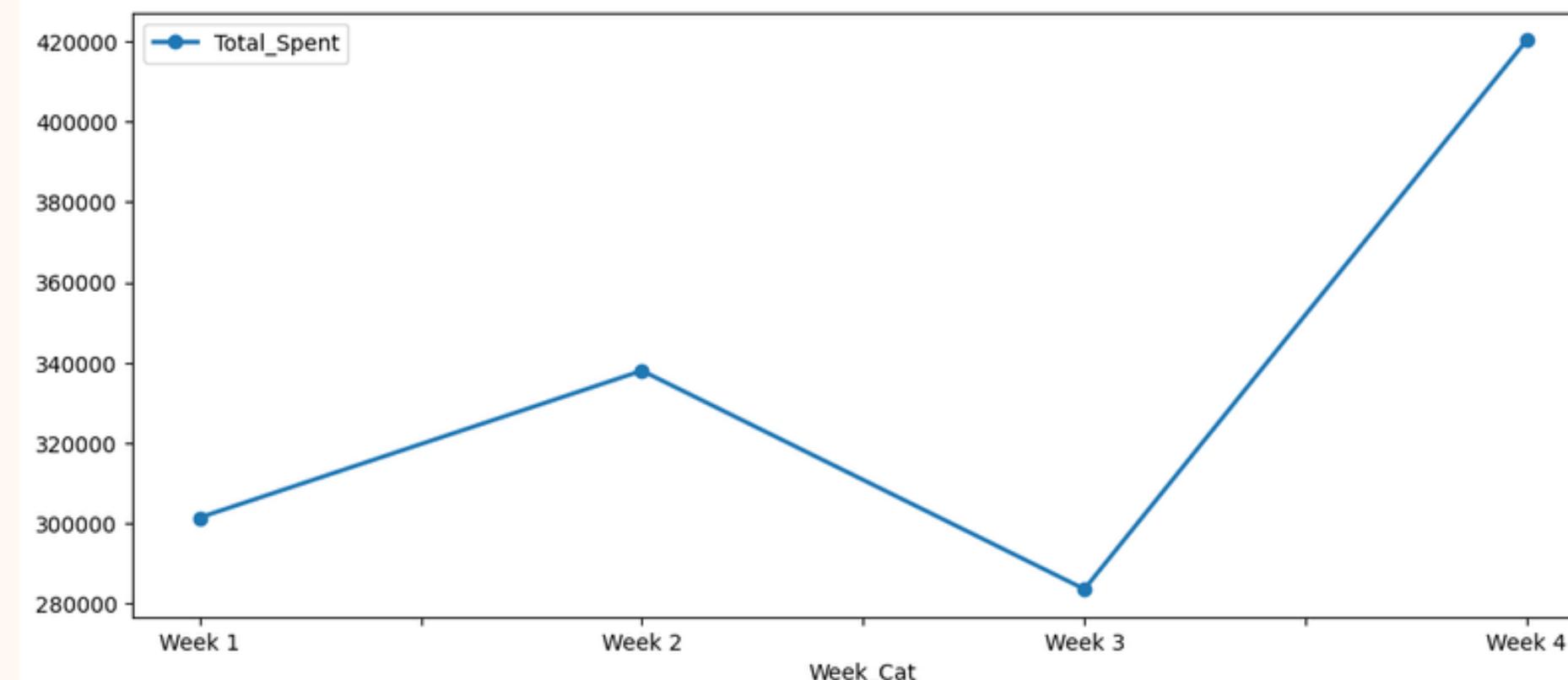


- Income dan Total Spent memiliki korelasi yang positif satu sama lain.
- Jumlah Total Spent yang ada tidak jauh berbeda untuk customer dengan range Income 0 sampai 60.000, perbedaan yang signifikan dapat terlihat pada customer dengan range income diatas 60.000 hingga 100.000.

TREND NOMINAL PEMBELANJAAN SECARA MINGGUAN DAN BULANAN



Data diatas menunjukan tren nominal pembelanjaan yang terus meningkat pada tahun 2012 sampai awal tahun 2013, setelahnya tren mengalami penurunan yang tajam hingga akhir tahun 2014, hingga nominal pembelanjaannya hampir kembali ke awal tahun 2012.



Secara Tren mingguan untuk nominal pembelanjaan dari week 1 ke week 2 mengalami peningkatan, namun menurun lagi secara tren bahkan ke level dibawah week 1. Tetapi pada week 3 secara melesat meningkat hampir 30% dalam 1 minggu. Uptrend line yang terdapat di week 3 terjadi karena kita asumsikan tanggal gajian orang-orang tepat di tanggal 25 atau week 3, disitulah terjadi pembelanjaan secara masif oleh para customer.

MODELLING



SPLIT DATA

Data displit dengan persentase 70% untuk training data dan 30% untuk test data

LABEL ENCODING

- age_cat
- Recency_cat
- Income_cat
- NumDeals_freq
- NumWeb_freq
- NumCat_Freq
- NumStr_Freq
- TotSpent_Cat
- Week_Cat

ONE HOT ENCODING

- Education
- Marital_Status



HYPERPARAMETER TUNING

MODEL	PRECISION	RECALL	F1-SCORE
KNN	0.34	0.34	0.34
LOGISTIC REGRESSION	0.59	0.34	0.43
DECISION TREE	0.41	0.42	0.42
RANDOM FOREST	0.46	0.34	0.39
XGBOOST	0.62	0.48	0.54



OVERSAMPLING (SMOTE)

MODEL	PRECISION	RECALL	F1-SCORE
KNN	0.26	0.46	0.34
LOGISTIC REGRESSION	0.42	0.48	0.45
DECISION TREE	0.39	0.55	0.45
RANDOM FOREST	0.52	0.56	0.54
XGBOOST	0.55	0.51	0.53

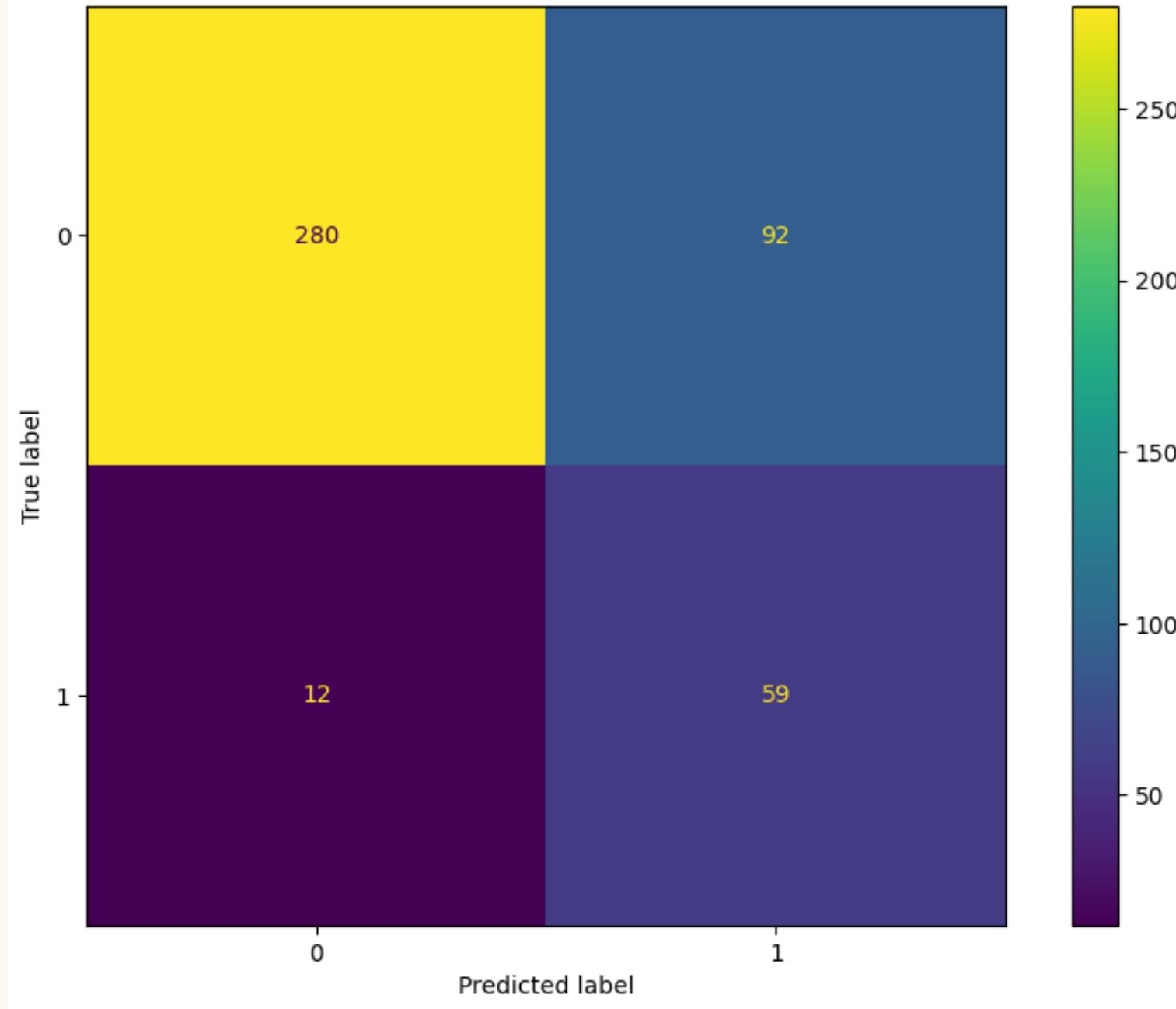


UNDERSAMPLING

MODEL	PRECISION	RECALL	F1-SCORE
KNN	0.27	0.66	0.39
LOGISTIC REGRESSION	0.42	0.48	0.45
DECISION TREE	0.31	0.75	0.43
RANDOM FOREST	0.35	0.80	0.49
XGBOOST	0.39	0.83	0.53

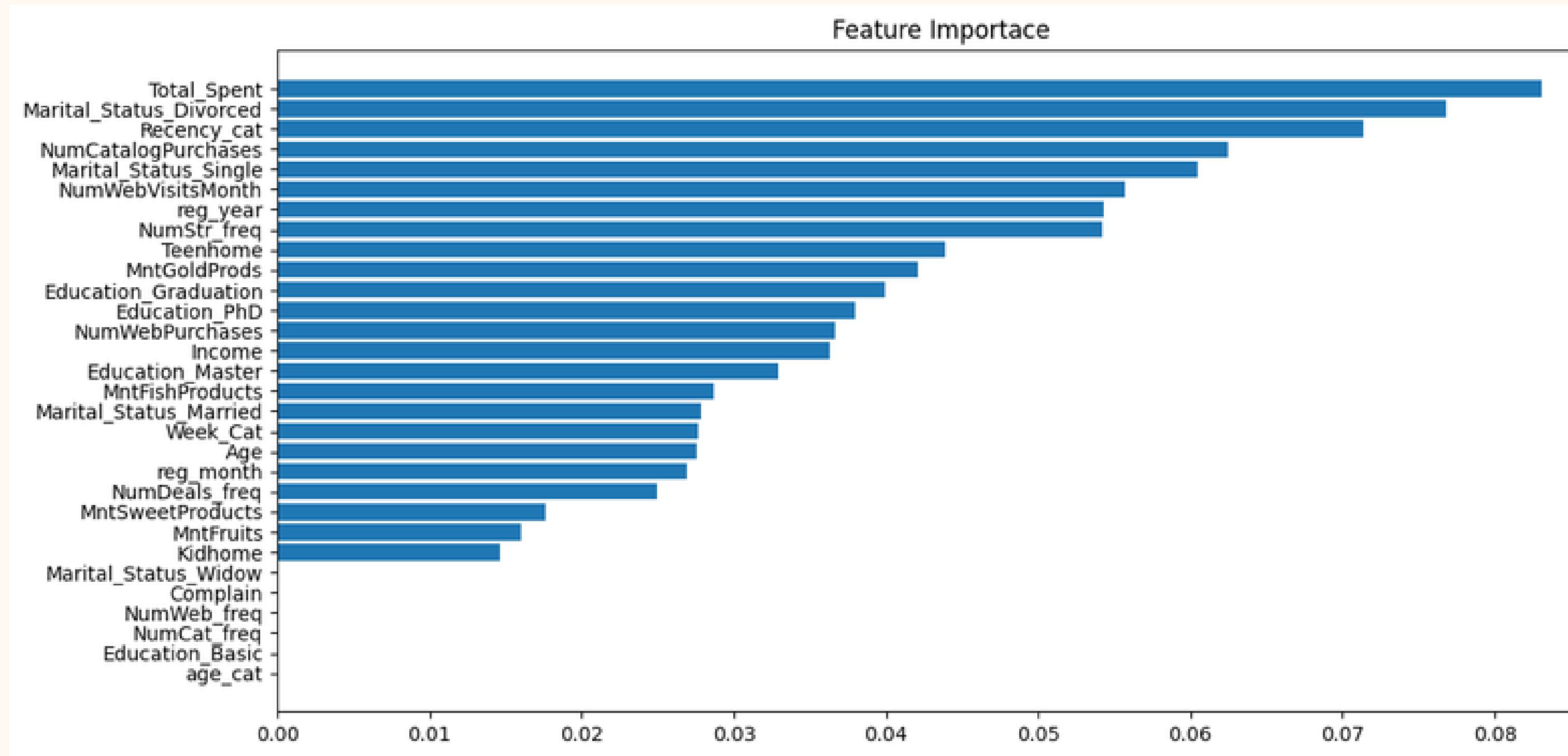


CONFUSION MATRIX



- Prediksi membeli membership dan actualnya membeli membership 59
- Prediksi tidak membeli membership dan actualnya membeli membership 12
- Prediksi tidak membeli memberhip dan actualnya tidak membeli membership 280
- Prediksi membeli membership dan actualnya tidak membeli memberhip 92

FEATURE IMPORTANCE



REKOMENDASI

Rekomendasi akan dibuat berdasarkan Feature Importance dari Model yang dipilih Jenis Customer yang perlu dihubungi untuk campaign selanjutnya

- Pelanggan dengan total pembelanjaan diatas 1200 USD
- Pelanggan dengan jangka waktu pembelian di range 0 - 20 hari dari tanggal pembelian sebelumnya
- Pelanggan dengan status pernikahan single dan divorced
- Pelanggan yang melakukan pembelian lebih dari 10 kali lewat Katalog,Web,dan Store
- Pelanggan yang tidak memiliki anak kecil maupun anak dewasa
- Pelanggan yang melakukan pembelanjaan di week ke-3

Rekomendasi tambahan

- Pertahankan kualitas produk yang dijual,karena semakin banyak customer yang membeli produk,semakin banyak juga persentase customer untuk membeli membership
- Menambah jenis produk yang ada,khusunya untuk anak kecil hingga dewasa karena secara persentase banyak customer yang membeli membership itu tidak memiliki anak kecil dan dewasa. Sehingga segmen market untuk anak kecil dan dewasa masih terbuka luas.
- Tingkatkan performa web,design katalog dan pelayanan di toko karena terdapat banyak customer yang melakukan pembelian lewat marketing channel tersebut.

DAMPAK TERHADAP BISNIS

Kita asumsikan pihak superstore belum melakukan tindakan lebih lanjut terkait strategi penjualan membership

- Harga satuan gold membership : 449 USD
- error rate tanpa permodelan : 70%
- Total customer yang membeli membership : 443

Dengan adanya permodelan

- Harga satuan gold membership : 449 USD
- error rate tanpa permodelan : 47%
- Total customer yang membeli membership : 443

Terdapat penurunan rate error 23% Total penjualan gold membership tanpa model $30\% \times 443 \times 449 = 59.672$ USD

Total penjualan gold membership dengan model $53\% \times 443 \times 449 = 105.420$ USD

Terjadi kenaikan revenue sebesar 45.748 USD