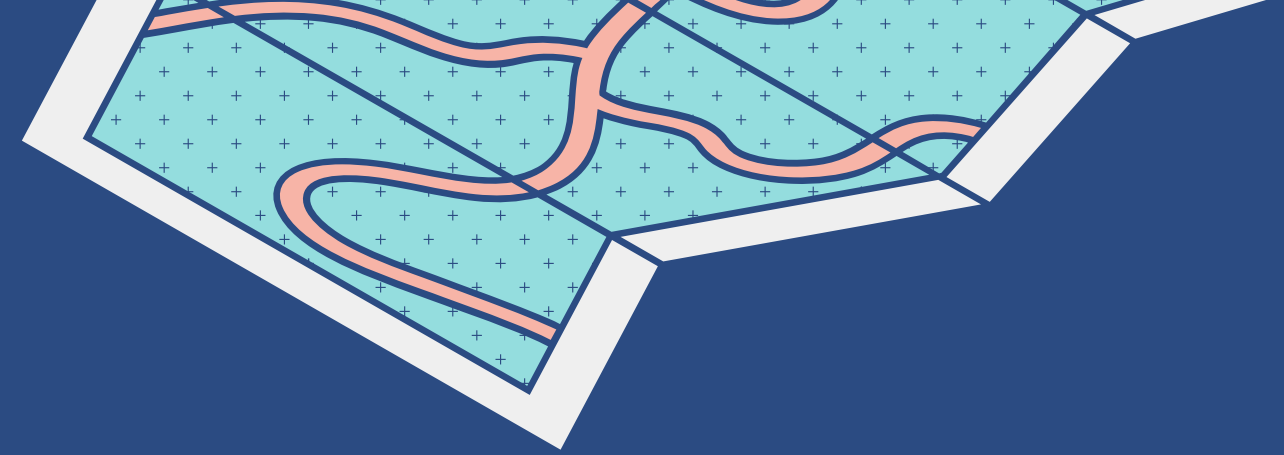




Marketing Campaign

Statstronauts



Our Company at Glance

PT. Statstronauts adalah perusahaan yang bergerak di bidang jasa marketing consultant khusus untuk digital marketing dengan memprioritaskan optimalisasi pemasaran.



MEET OUR TEAM



David



Nabila



Morrena



Jakfar



Ivan



Prajaka



Idhar



Havis



Kak Kevin

Table Of Content

- Business Understanding
- EDA
- Data Preprocessing
- Modelling & Evaluation
- Business Recommendation
- Business Simulation

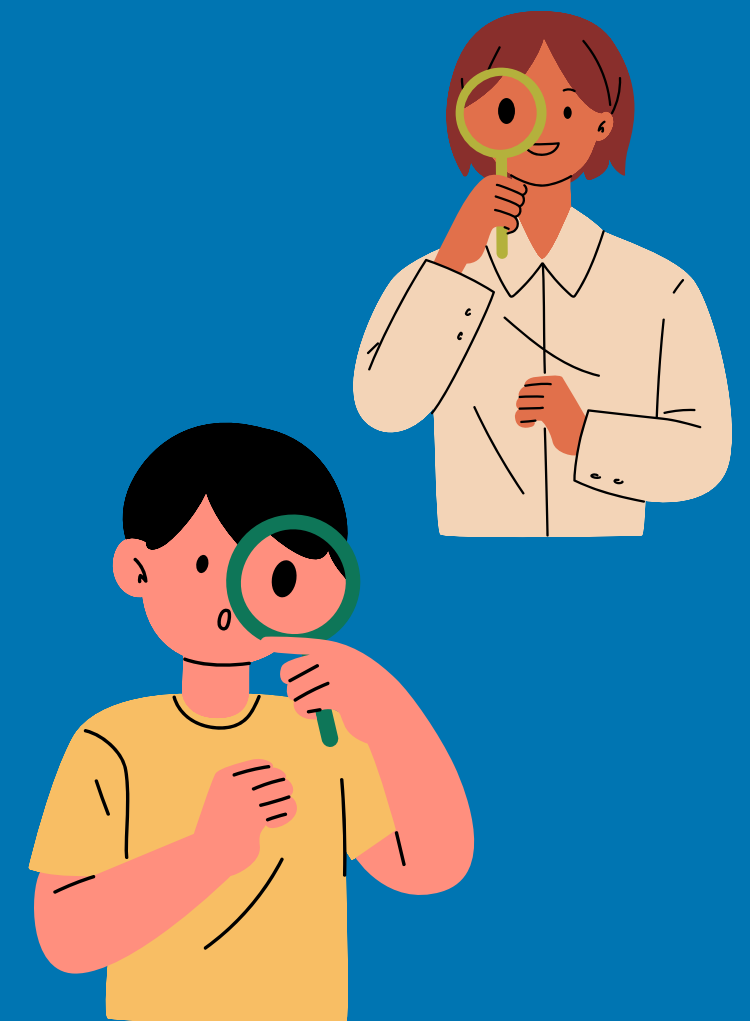





Business Understanding

Problems


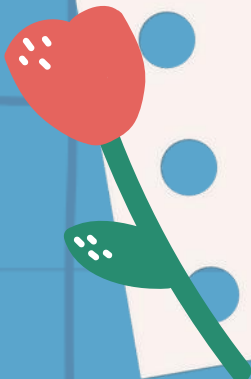
- Lack of Data dari customer yang berpotensi menerima campaign yang ditawarkan selanjutnya.
- Pengefektifan strategi pemasaran.
- Meningkatkan rate keberhasilan campaign selanjutnya.




Goals & Objective



Meningkatkan
conversion rate dari
marketing campaign
agar strategi
pemasaran yang
dilakukan sukses.



Membuat suatu model
prediksi yang dapat
mengklasifikasikan
pengguna yang
berpotensi menerima
campaign yang
ditawarkan.





Business Metrics

CONVERSION RATE%

adalah persentase yang biasanya digunakan dalam pemasaran digital untuk mengevaluasi kinerja lalu lintas situs web, kampanye pemasaran, dan konversi.

$$\text{Conversion Rate} = \frac{\text{Total Conversions}}{\text{Total Visitors}}$$



Exploratory Data Analysis

Descriptive Analysis

Terdapat data umur yang aneh, yaitu min 1893 Income max nya 666666, dan jauh dari mean (52427) untuk Marital Status ada beberapa kategori yang bisa digabung agar data menjadi lebih sederhana, Together dan Married, Widow dan Divorced untuk Education ada beberapa kategori 2n Cycle bisa digabung dengan Master karena sama

```
df[nums].describe()
```

	Income
count	2216.000000
mean	52247.251354
max	666666.000000

Descriptive Analysis

- Terdapat 24 data kosong pada kolom data income
- Tidak adanya data yang duplicate

```
df_duplicate = df[df.duplicated(keep = 'last')]  
df_duplicate
```

```
ID  Year_Birth  Education  Marital_Status  In
```

```
0 rows x 29 columns
```

```
df.isna().sum()
```

ID	0
Year_Birth	0
Education	0
Marital_Status	0
Income	24
Kidhome	0
Teenhome	0
Dt_Customer	0

Descriptive Analysis

- Biaya Campaign per callnya adalah 3\$
- Revenue yang didapatkan jika campaign diterima adalah 11\$

```
[ ] df['Z_CostContact'].unique()
```

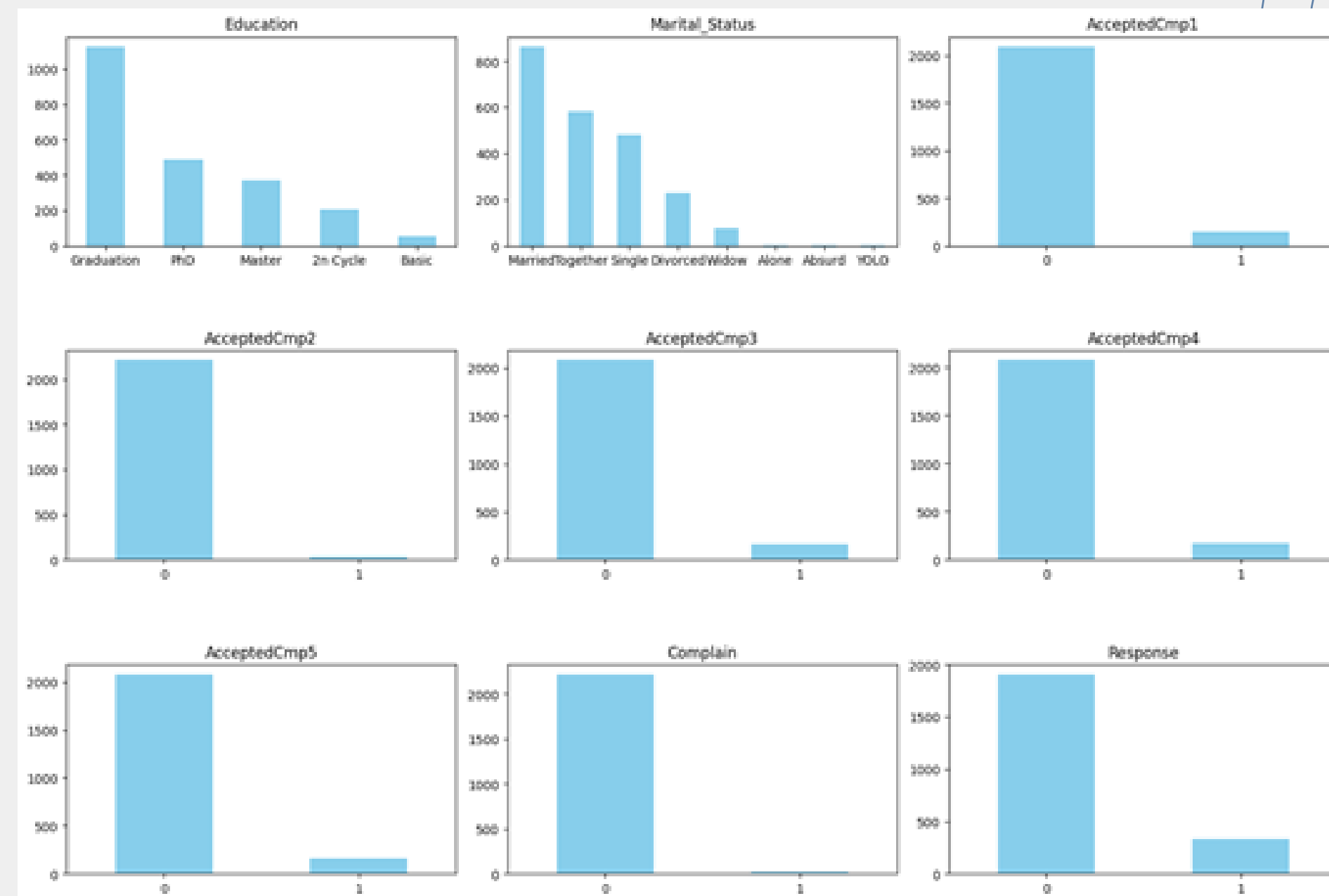
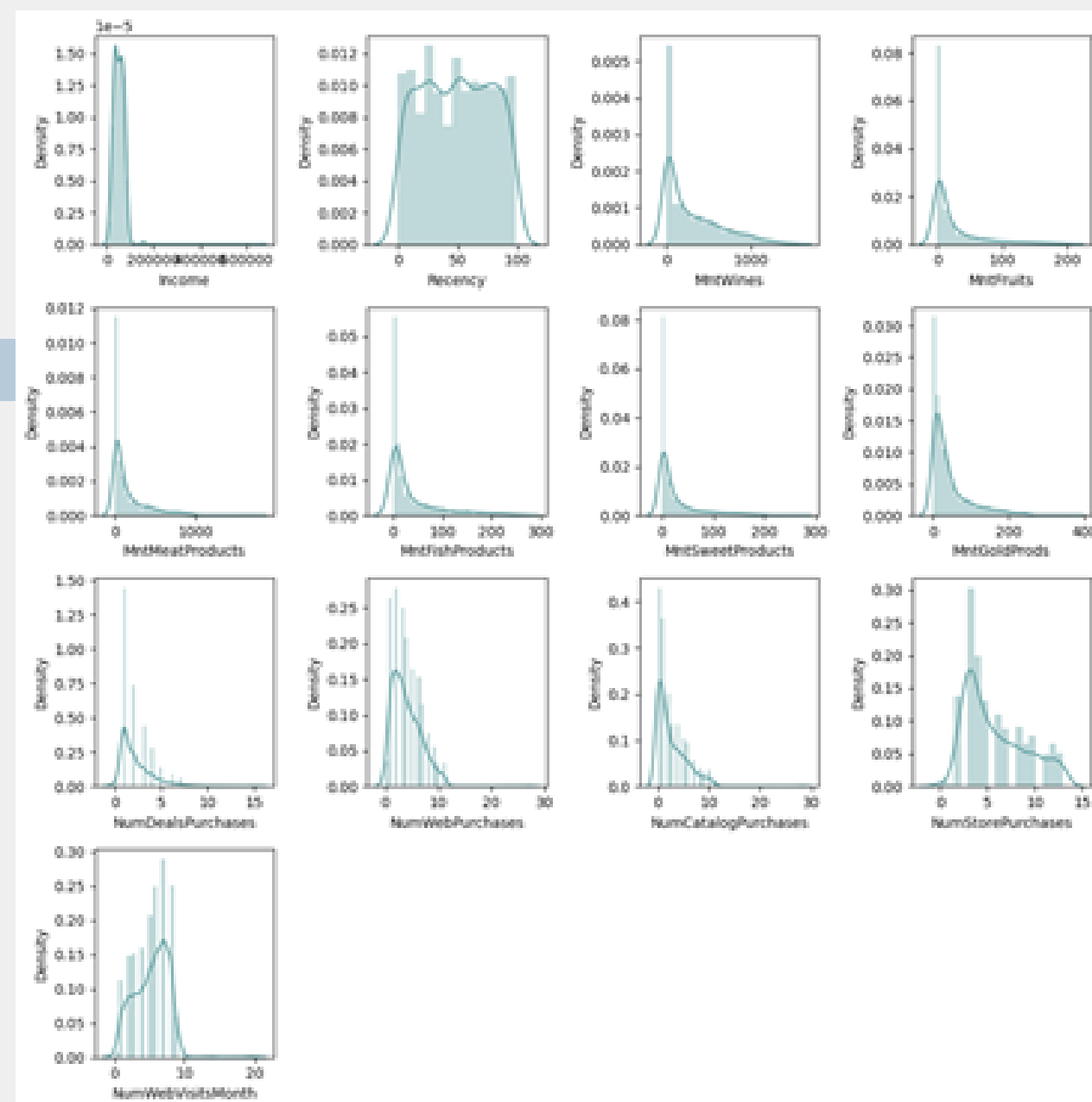
```
array([3])
```

```
[ ] df['Z_Revenue'].unique()
```

```
array([11])
```



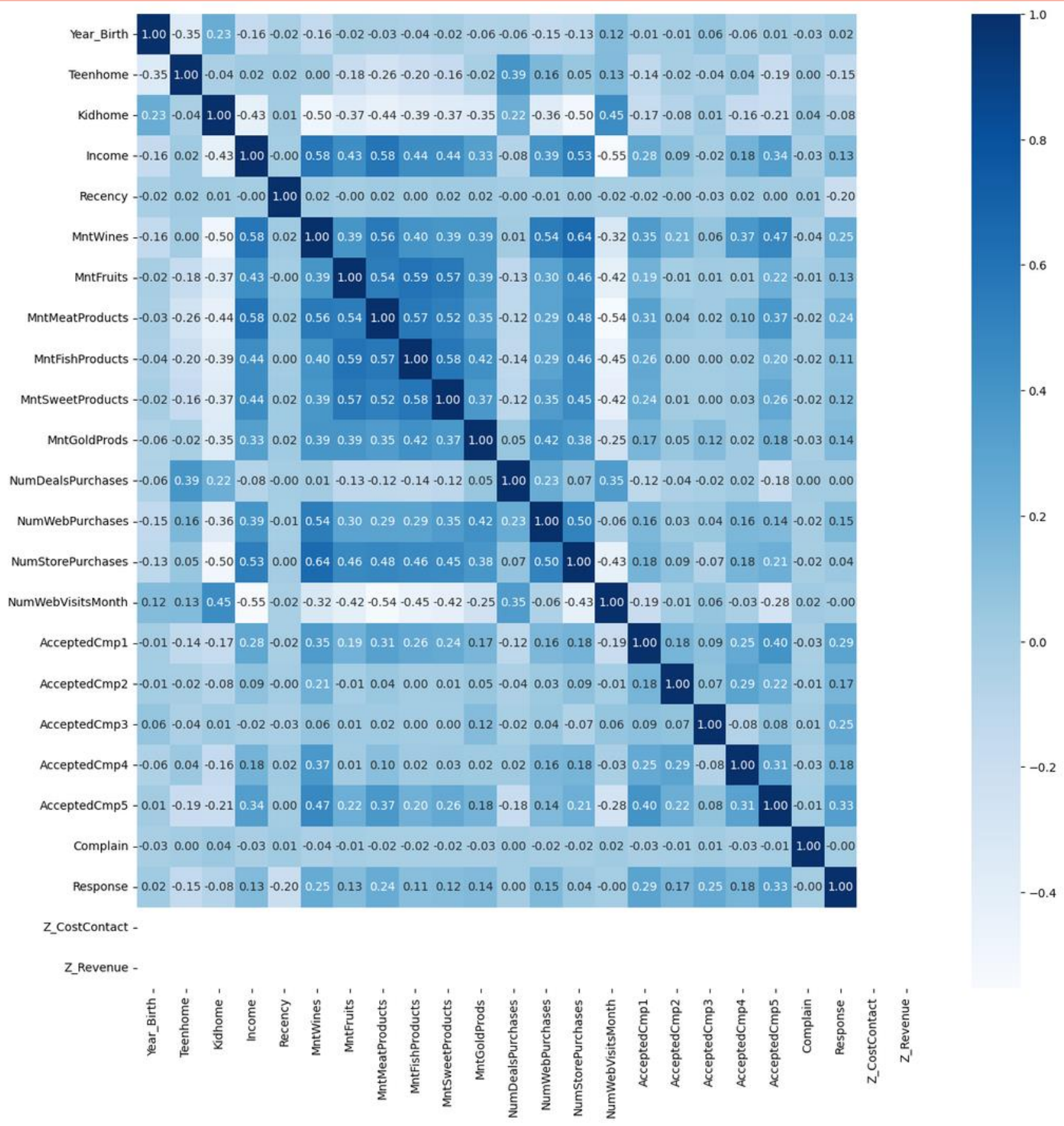
Univariate Analysis



Positively Skewed : Income, Amount of Wines, Fruit, Meat, Fish, Sweet, Gold, Deals Purchase, Web Purchase, Catalog Purchase, Store Purchase
Negatively Skewed : Web Visit per Month

Multivariate Analysis

OVERVIEW



Campaign ke 5 merupakan campaign dengan korelasi tertinggi pada Response
Jenis campaign ke 5 merupakan yang paling efektif untuk menarik respons.



Wine merupakan produk dengan korelasi tertinggi pada Response
Perlunya penyediaan stock wines pada offering



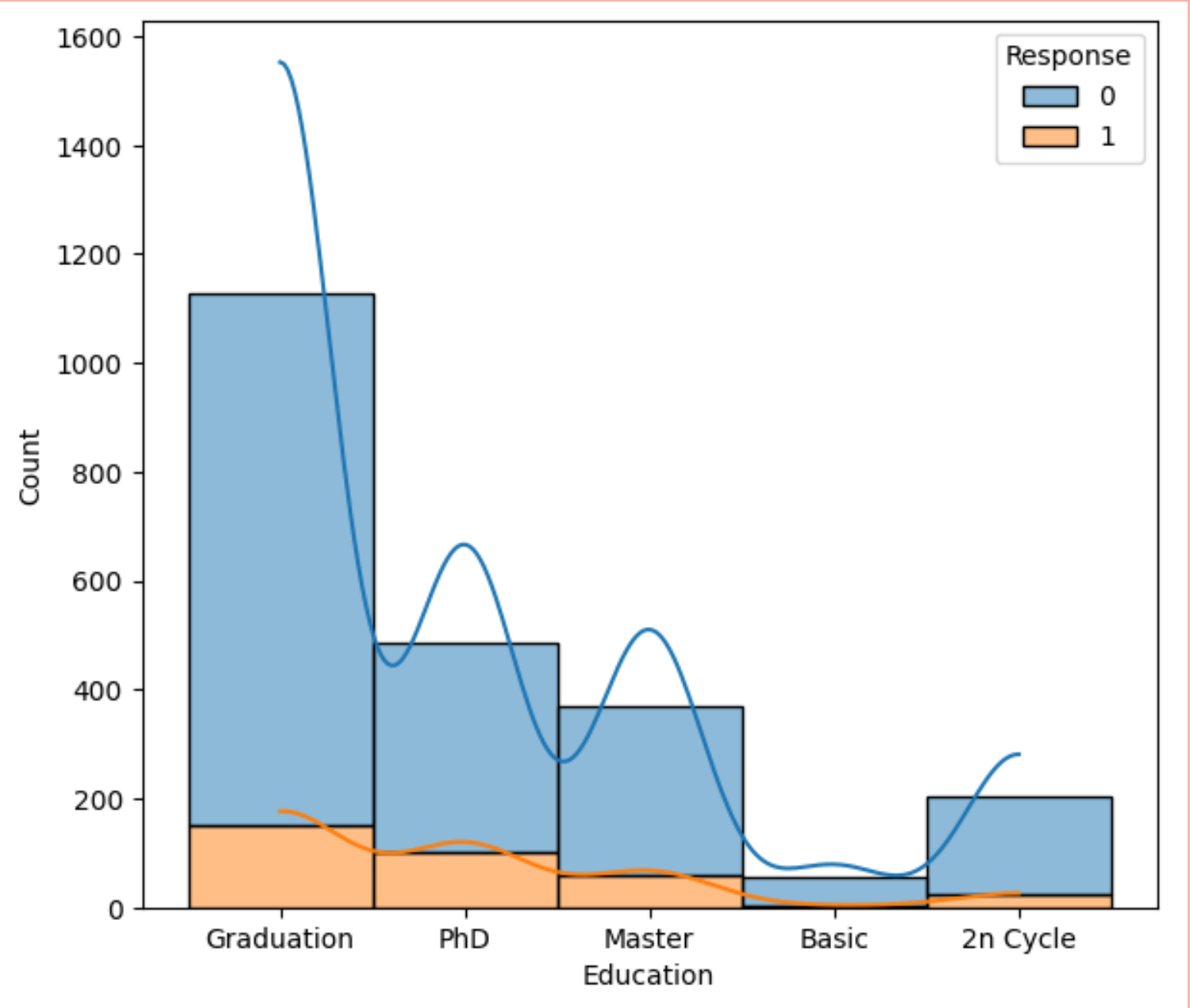
Number of Web Purchase merupakan tipe pembelian dengan korelasi tertinggi pada respons
WebPurchase bisa menjadi pilihan utama platform untuk next campaign



Jumlah Anak Kecil dan Remaja di rumah memiliki korelasi negatif terhadap respons
Tidak efektif memberikan campaign pada keluarga yang sudah memiliki anak

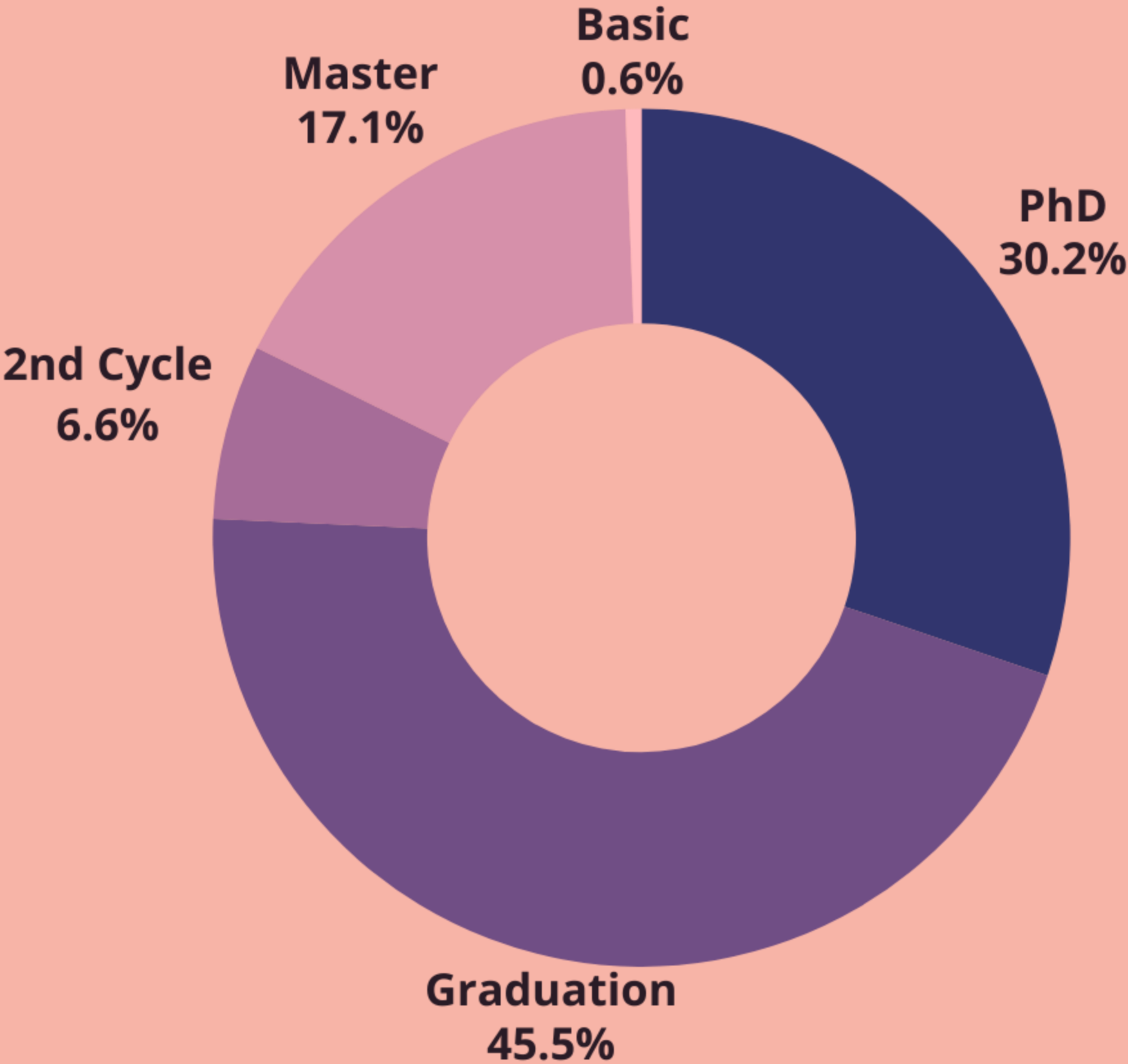
Multivariate Analysis

EDUCATIONAL CATEGORICAL DATA



EDUCATION

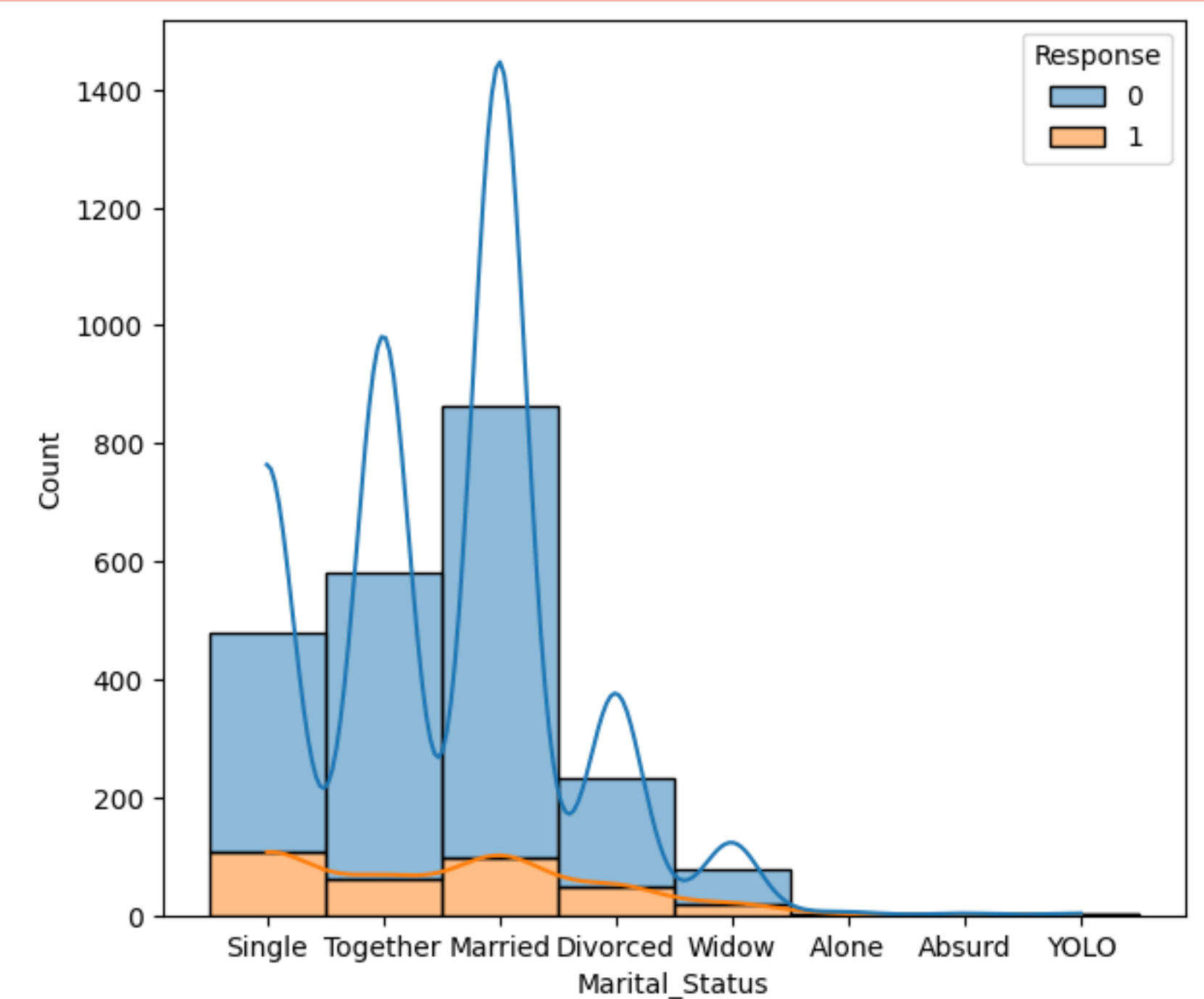
Customer dengan Graduation Degree merupakan responden paling besar terhadap campaign



Total Response for Accepted Campaign Based on Education

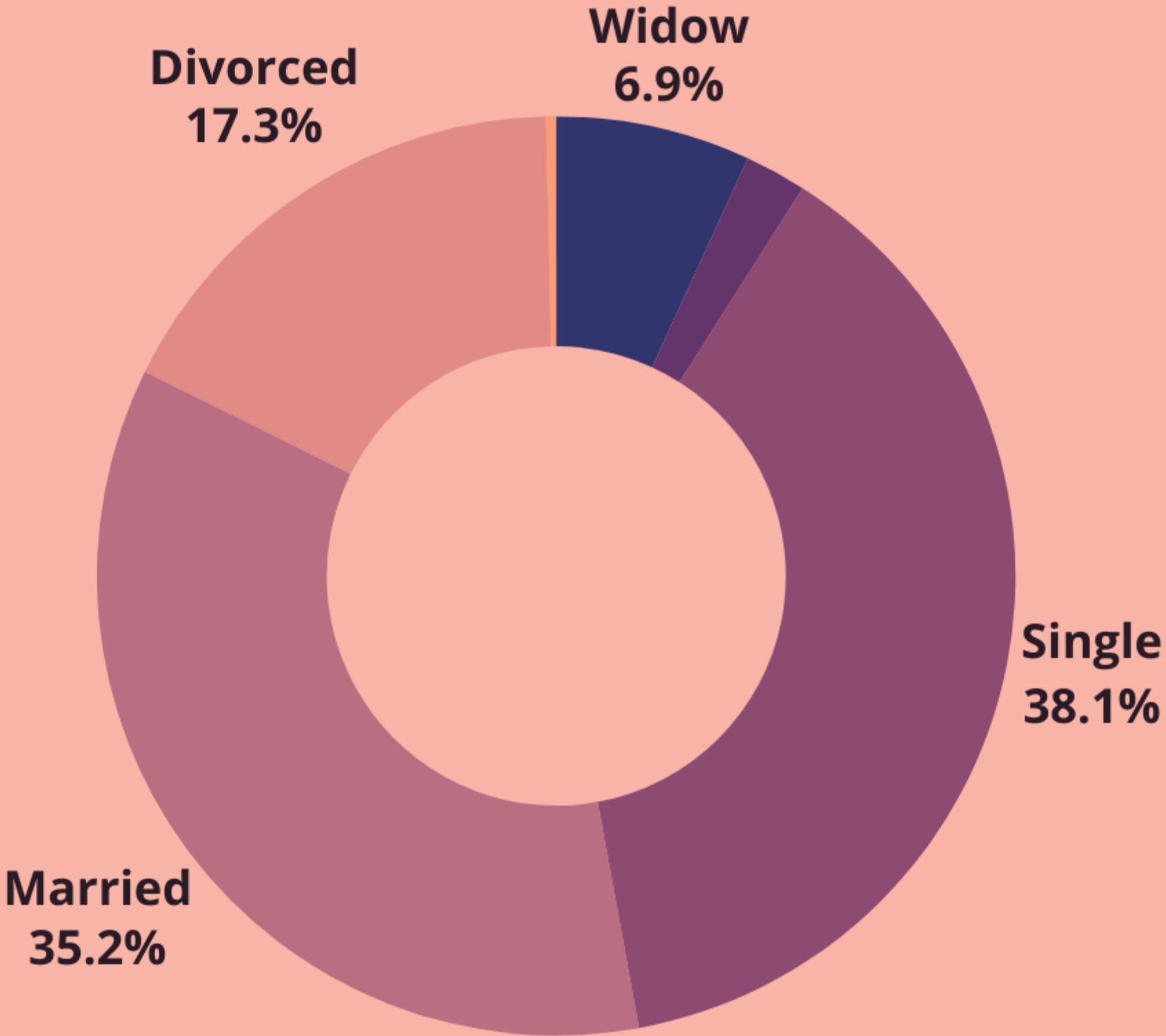
Multivariate Analysis

CATEGORICAL DATA



MARITAL STATUS

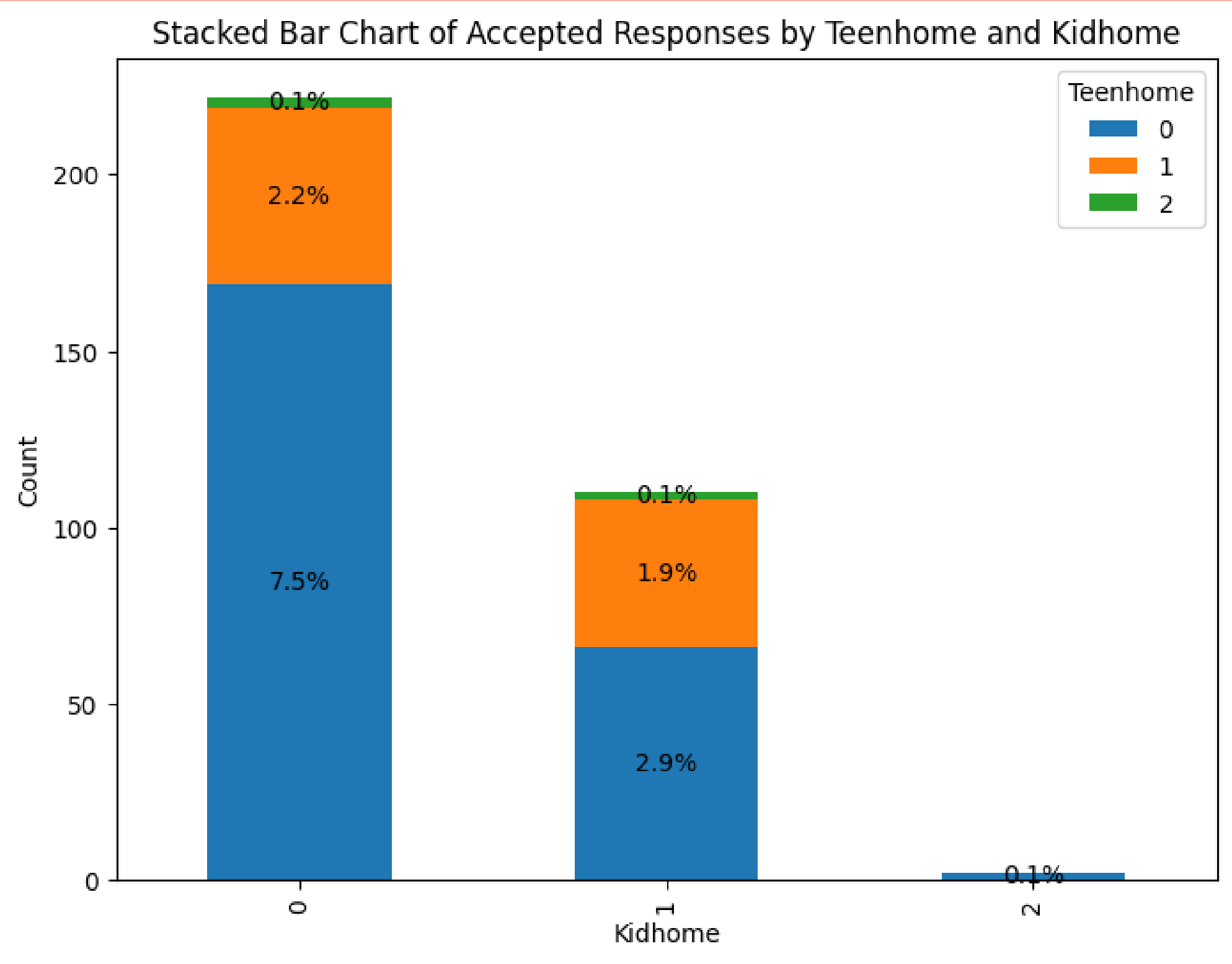
Customer dengan status Single merupakan responden paling besar terhadap campaign



Total Response for Accepted Campaign
Based on Marital Status

Multivariate Analysis

CATEGORICAL DATA



NUMBER OF KIDS AND TEENS

Customer dengan accepted Campaign terbanyak merupakan customer dengan jumlah 0 Teen dan Kid.



Data Pre-Processing

Data Cleaning

Handling Missing Value

Menghapus missing values (1,07 % dari total data)

Handling Duplicate Data

Tidak ada baris yang duplicate

Handling Outlier

Menggunakan Z-Score untuk mendeteksi dan menghapus outliers. Data yang terhapus: 10,07% dari data train

Handling Invalid Data

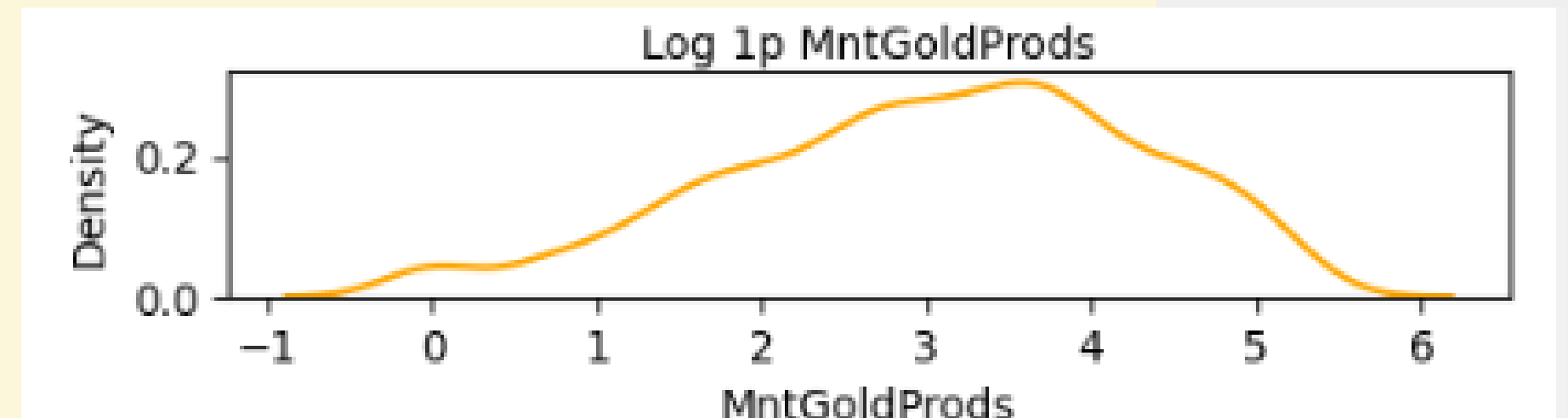
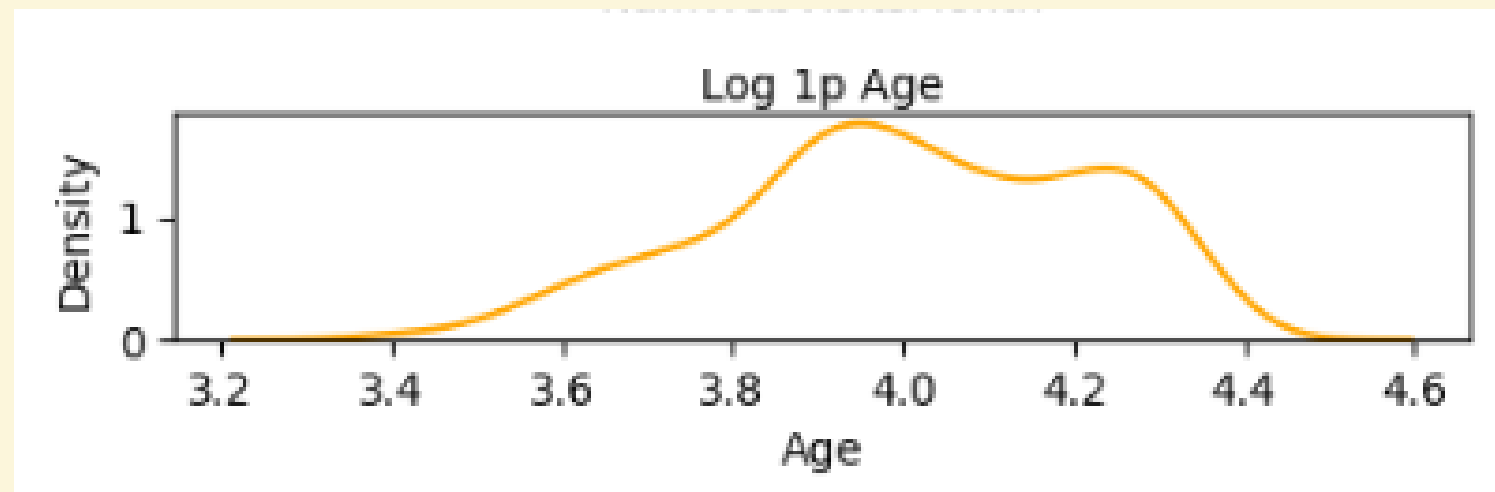
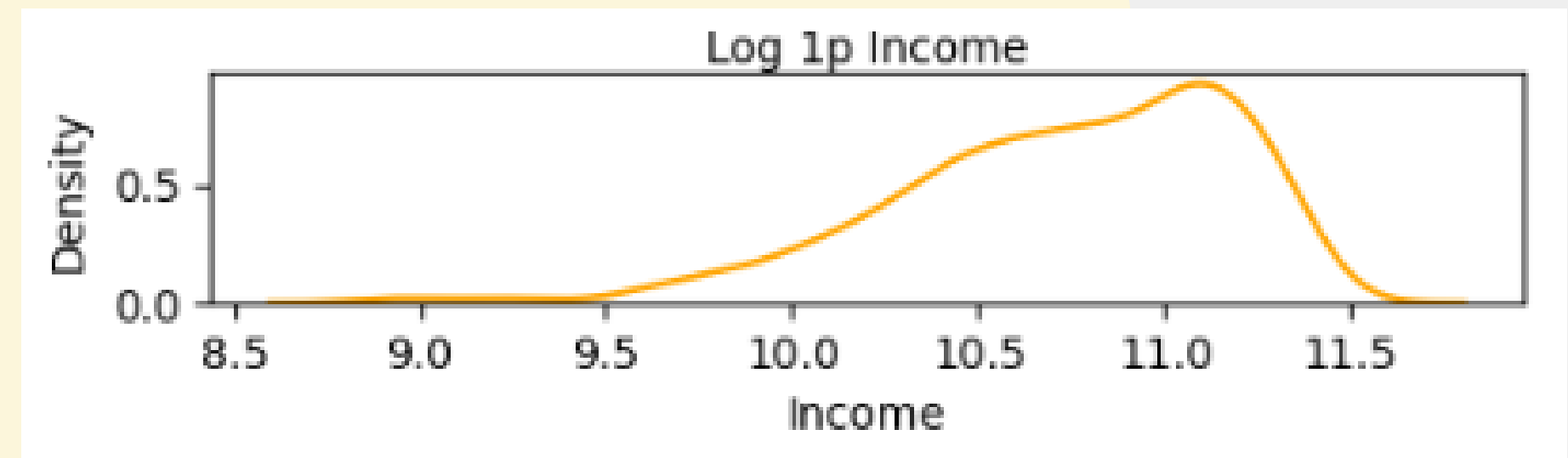
Menghapus absurd, dan YOLO (0,17% dari total data)



Feature Transformation

Semua fitur yang memiliki tipe data numeric melakukan Feature Transformation menggunakan $\log + 1$ hasilnya hanya beberapa fitur yang mendekati distribusi normal setelah dilakukan Feature Transformation :

- Income
- MntGoldProds
- Age



Feature Extraction

* _____ *				
* _____ *				
* _____ *				
* _____ *				
* _____ *				
Age	Varian_products	Total_products	TotalAccepted	TotalPurchases
Membuat Feature Age berdasarkan feature Year Birth dikurang dengan tahun sekarang	Menghitung berapa banyak variasi produk yang dibeli oleh customer	Menghitung berapa banyak produk yang dibeli oleh customer	Menghitung berapa kali customer pernah menerima campaign	Menghitung berapa kali customer berbelanja lewat platform apapun

Feature Encoding

Menggunakan Label Encoding untuk mengubah Feature Categorical seperti Marital_Status, dan Education menjadi numeric agar dapat menjadi input untuk model machine learning

Feature Engineering

Menerapkan standardisasi terhadap kolom numerik yang telah dilakukan log transformation.

Feature Selection

Numerical Feature

Menggunakan SelectKBest dengan score function `f_classif` untuk mengurutkan feature numerik dari yang paling penting ke yang tidak terlalu penting

Categorical Feature

Menggunakan SelectKBest dengan score function `chi square` untuk mengurutkan feature kategorikal dari yang paling penting ke yang tidak terlalu penting

****Notes**


Pemilihan feature dilakukan dengan hati-hati dengan memperhatikan feature agar tidak redundant

	Feature	Score
24	TotalAccepted	221.903526
18	AcceptedCmp1	130.648565
17	AcceptedCmp5	98.394283
15	AcceptedCmp3	74.621768
12	NumCatalogPurchases	72.462120
6	MntMeatProducts	50.796009
3	Recency	49.372085
9	MntGoldProds	48.349267
23	Total_products	46.898832
16	AcceptedCmp4	44.165055
19	AcceptedCmp2	43.451439
11	NumWebPurchases	35.035350
25	TotalPurchases	30.452916
2	Teenhome	28.825366
4	MntWines	26.369837
8	MntSweetProducts	22.809607
5	MntFruits	16.835848
0	Income	12.093496
7	MntFishProducts	11.181152
1	Kidhome	5.314368
22	Varian_products	4.595611
20	Complain	0.341484
14	NumWebVisitsMonth	0.238056
13	NumStorePurchases	0.100034
10	NumDealsPurchases	0.004716
21	Age	0.000973

Feature Selection

Feature yang diambil :

- **TotalAccepted** : Total Campaign diterima
- **Recency** : Jumlah hari sejak pembelian terakhir
- **Teenhome** : Jumlah anak remaja di rumah
- **Total_products** : Jumlah produk yang pernah dibeli oleh customer
- **Education** : Tingkat pendidikan customer
- **Marital_together** : Indikator apakah customer tinggal dengan orang lain
- **Marital_single** : Indikator apakah customer tinggal sendirian
- **TotalPurchases** : Jumlah berapa kali customer berbelanja lewat platform apapun

29 Fitur  **8 Fitur**

Class Imbalance

Numerical Feature

Menggunakan Oversampling SMOTE pada data train untuk handle class imbalance menjadi perbandingan 50:50



Modeling and Evaluation

Model Evaluation Metrics

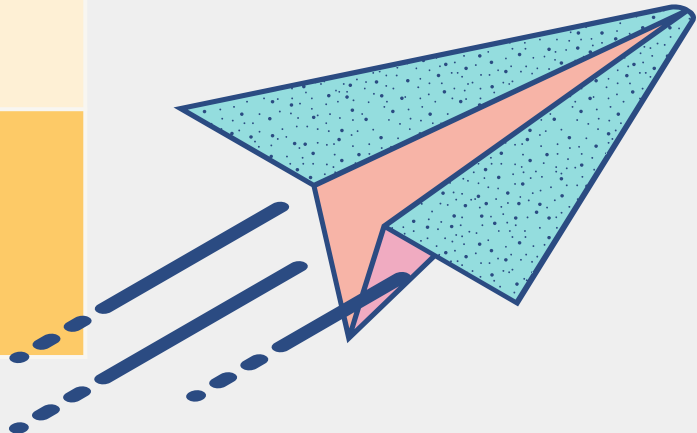
PRECISION

Precision dipilih sebagai metric evaluation menyesuaikan dengan objective project: yaitu mengurangi sebanyak mungkin false positive.



Modeling Results (Before Hypertuning)

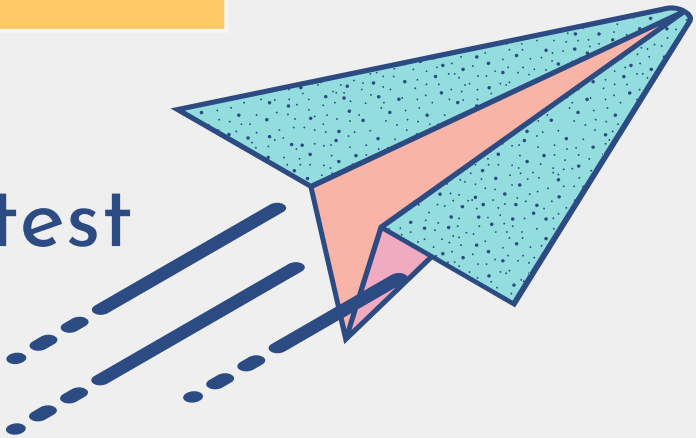
Model	Train Precision	Test Precision	Gap
Decision Tree	99,7%	71,4%	28,3%
Random Forest Classifier	99,7%	83,8%	15,9%
SVM	74,4%	85,2%	-10,8%
AdaBoost	99,7%	82,1%	17,6%
XGB	91,9%	82,7%	9.2%



Modeling Results (After Hypertuning)

Model	Train Precision	Test Precision	Gap
SVM	94,3%	76,1%	18,2%
Random Forest Classifier	99,6%	83,9%	15,7%
XGBoost	91,7%	85%	6,7%

Model yang dipilih adalah **XGBoost** dikarenakan hasil score nya yang paling tinggi dibanding yang lain, sekaligus gap antara data train dan test yang rendah (tidak overfitting dan underfitting)



Conversion Rate Improvement

	Sebelum Model	Sesudah Model
Call	2240	334
Reponse Positif	589	224



Conversion Rate Improvement

14.9%
Before Model

+23.13%



38.03%
After Model



Loss to Profit

-3.046\$

Before Model

+3.743\$



+697\$

Afer Model

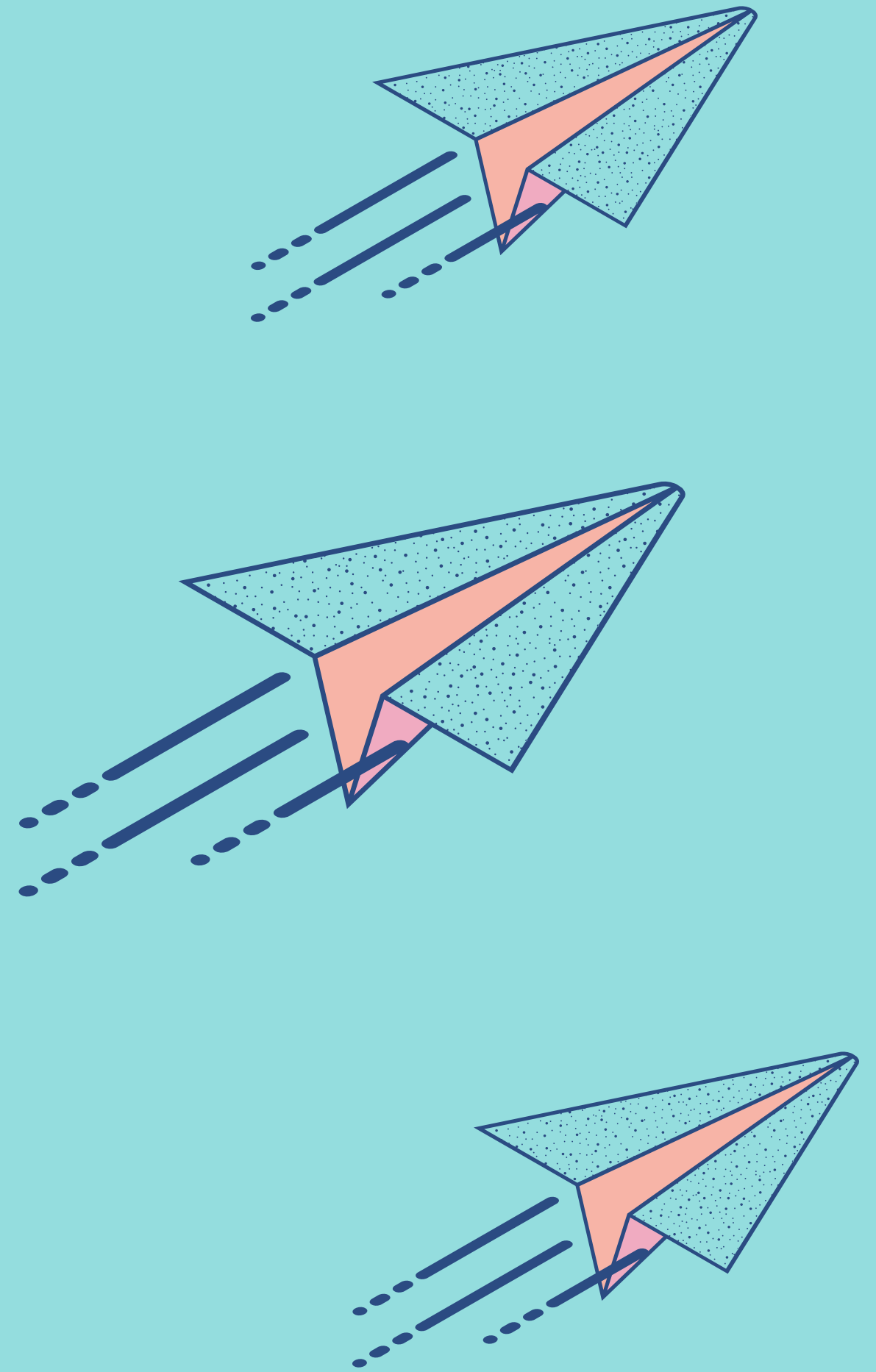
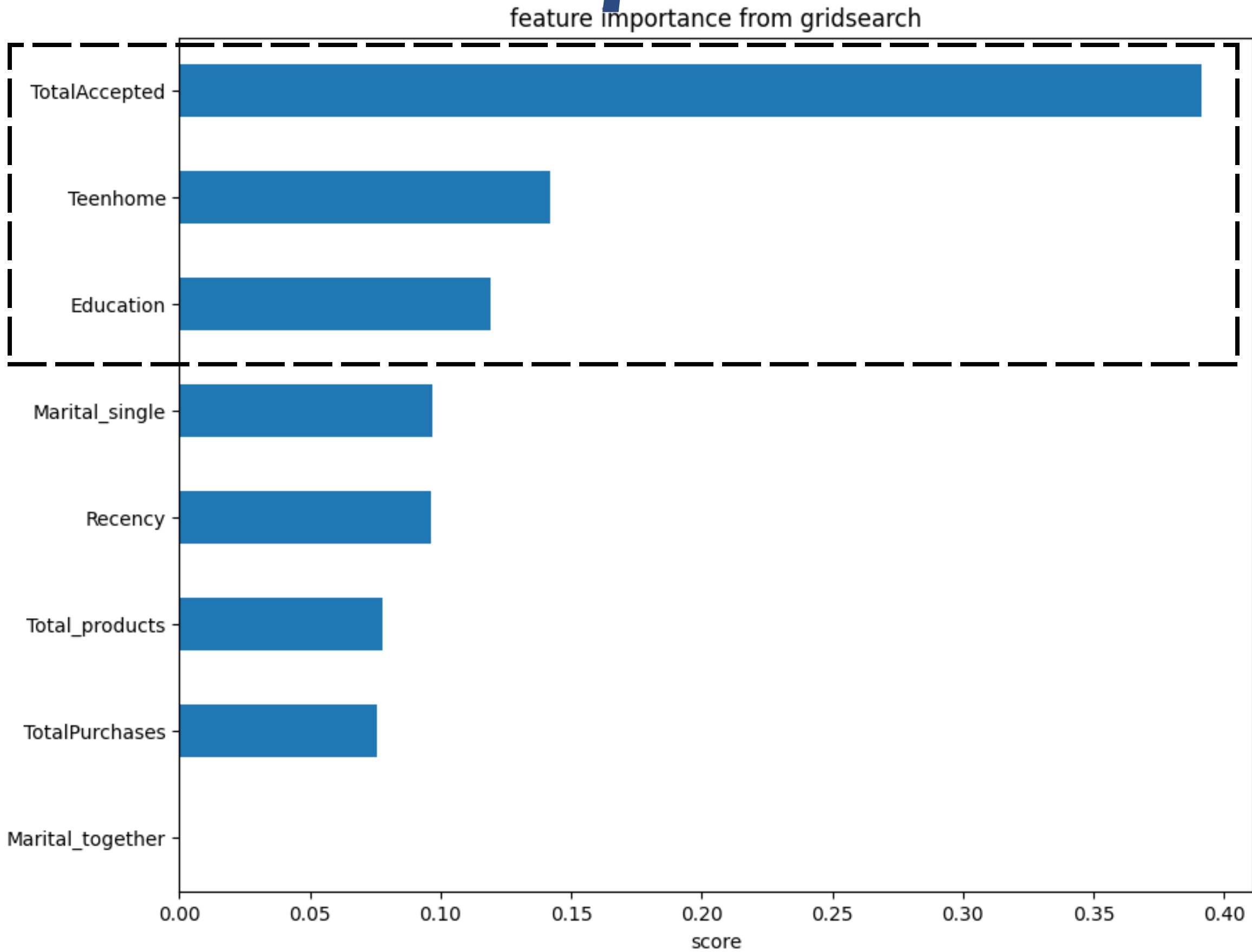
****Note :**

Cost call = 3\$/Call

Revenue = 11\$/Accept

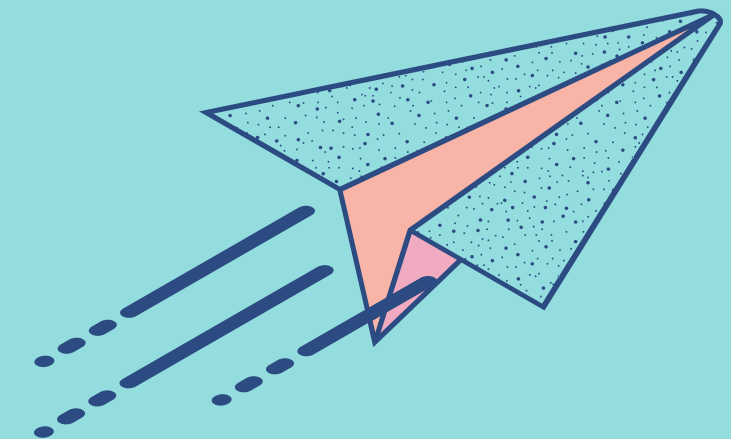


Feature Importance

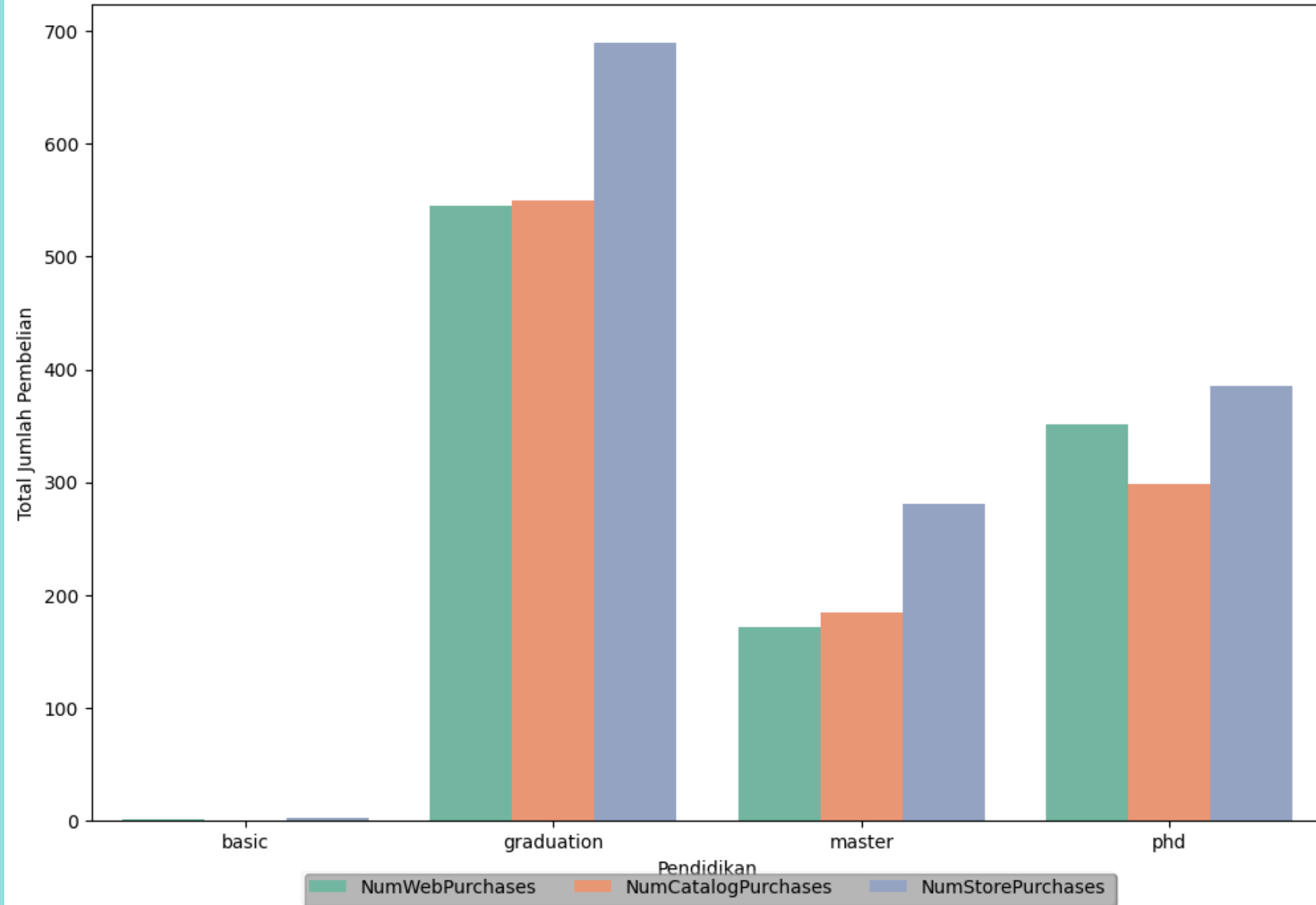


Feature Importance

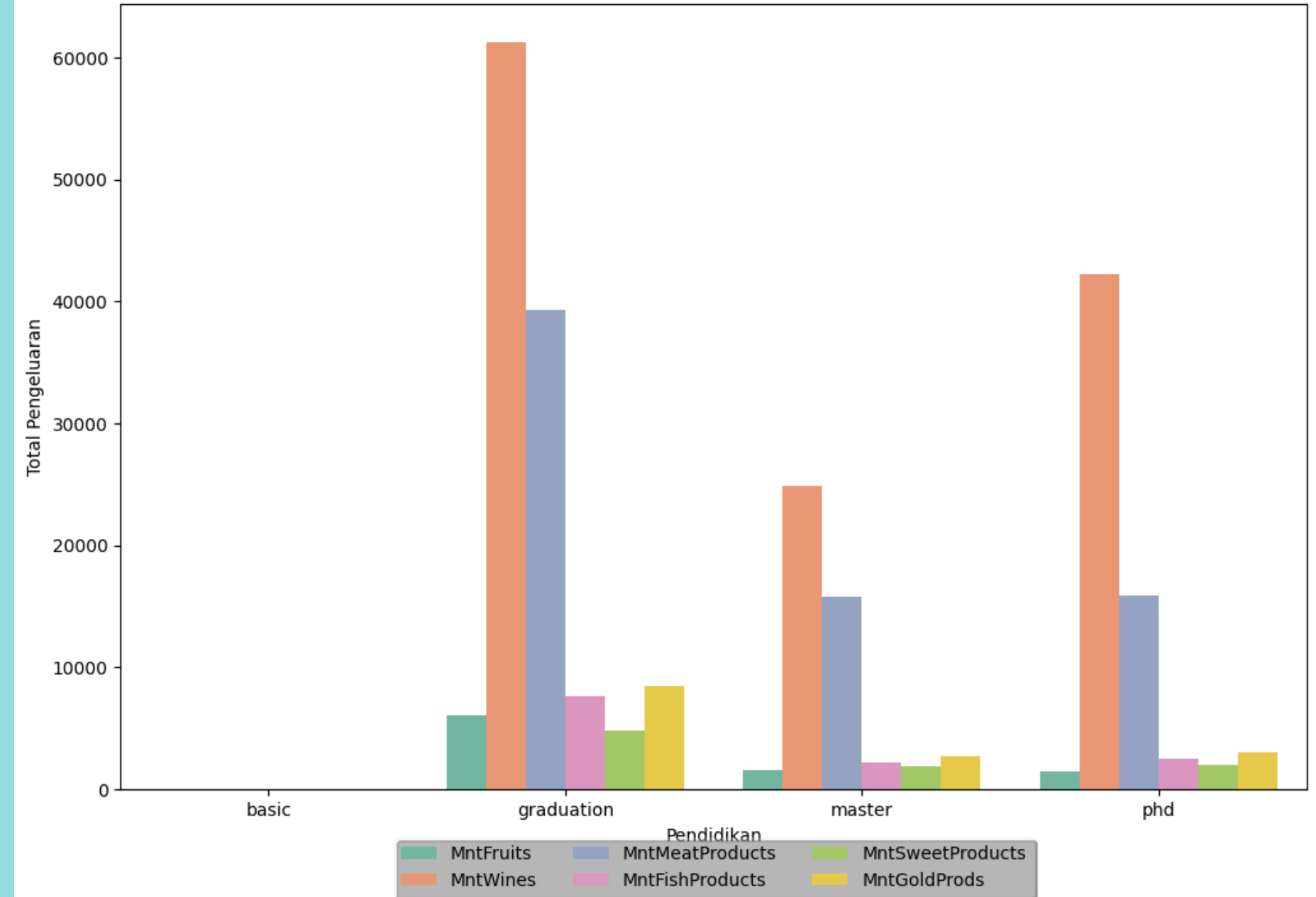
Pendidikan



Total Jumlah Pembelian berdasarkan Pendidikan



Total Pengeluaran pada Produk berdasarkan Pendidikan



Business Recommendation



Recommendation 1:

Campaign selanjutnya lebih mengutamakan customer yang sudah pernah accept campaign sebelumnya

Recommendation 2:

Dikarenakan Keberadaan Teenhome berada pada tingkat kedua, disarkankan membuat promo yang meningkatkan minat anak

Recommendation 3:

Dikarenakan Keberadaan Education berada pada tingkat ketiga disarkankan membuat promo yang menyasar Customer dengan tingkat pendidikan basic karena memiliki Response paling tinggi dibandingkan kelompok yang lain



**THANK
YOU**