

Marketing Campaign

Kelompok 2 LASKAR ONLINE

Dokumen
Laporan Final
Project



Nama Anggota Kelompok

Angelus Felix Sihombing

Aisyah Raudhatuzzahra

Saip Ardo Pratama

Edhita Kristasari

Richard Noel

Marha Nur Amalina

Andri Lau

NO 1

A. Apakah ada kolom dengan tipe data kurang sesuai, atau nama kolom dan isinya kurang sesuai?

- Kolom '**DT Customer**' bertipe string, seharusnya bertipe date
- Kolom **AcceptedCmp 1 - 5**, '**Complain**', dan '**Response**' bertipe integer, seharusnya bertipe boolean

B. Apakah ada kolom yang memiliki nilai kosong? Jika ada, apa saja?

Kolom '**income**' memiliki **24** baris data null

C. Apakah ada kolom yang memiliki nilai summary agak aneh?

(min/mean/median/max/unique/top/freq)

- Data yang mean dan median berbeda cukup jauh: '**Kidhome**', '**Teenhome**', '**MntWines**', '**MntFruits**', '**MntMeatProducts**', '**MntFishProducts**', '**MntSweetProducts**'
- Nilai max yang tidak normal: '**Income**'

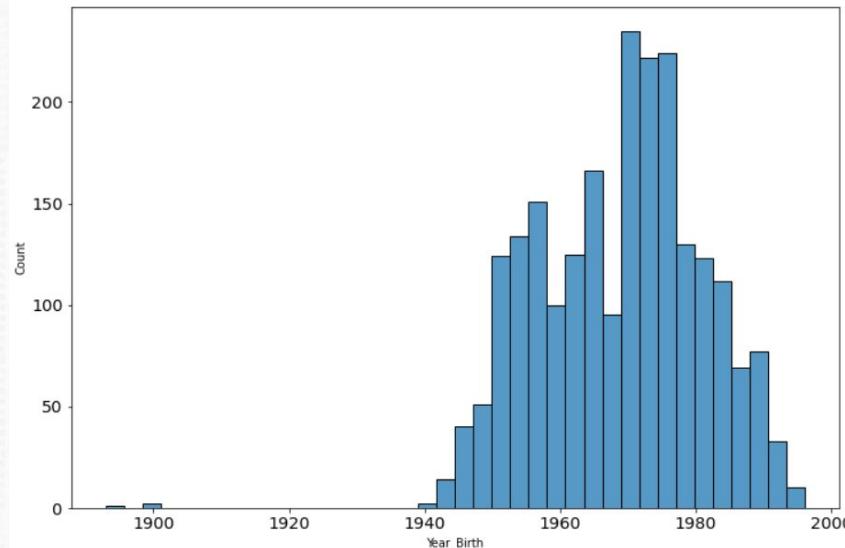
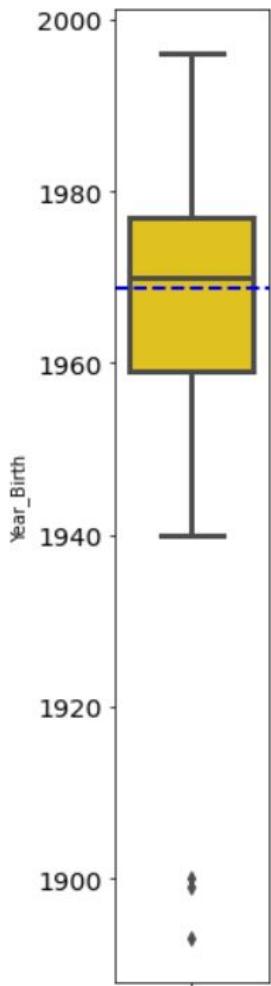
0	ID	2240	non-null	int64
1	Year_Birth	2240	non-null	int64
2	Education	2240	non-null	object
3	Marital_Status	2240	non-null	object
4	Income	2216	non-null	float64
5	Kidhome	2240	non-null	int64
6	Teenhome	2240	non-null	int64
7	Dt_Customer	2240	non-null	object
8	Recency	2240	non-null	int64
9	MntWines	2240	non-null	int64
10	MntFruits	2240	non-null	int64
11	MntMeatProducts	2240	non-null	int64
12	MntFishProducts	2240	non-null	int64
13	MntSweetProducts	2240	non-null	int64
14	MntGoldProds	2240	non-null	int64
15	NumDealsPurchases	2240	non-null	int64
16	NumWebPurchases	2240	non-null	int64
17	NumCatalogPurchases	2240	non-null	int64
18	NumStorePurchases	2240	non-null	int64
19	NumWebVisitsMonth	2240	non-null	int64
20	AcceptedCmp3	2240	non-null	int64
21	AcceptedCmp4	2240	non-null	int64
22	AcceptedCmp5	2240	non-null	int64
23	AcceptedCmp1	2240	non-null	int64
24	AcceptedCmp2	2240	non-null	int64
25	Complain	2240	non-null	int64
26	Z_CostContact	2240	non-null	int64
27	Z_Revenue	2240	non-null	int64
28	Response	2240	non-null	int64

		count	mean	std	min	25%	50%	75%	max
	Kidhome	2240.0	0.444196	0.538398	0.0	0.00	0.0	1.00	2.0
	Teenhome	2240.0	0.506250	0.544538	0.0	0.00	0.0	1.00	2.0
	MntWines	2240.0	303.935714	336.597393	0.0	23.75	173.5	504.25	1493.0
	MntFruits	2240.0	26.302232	39.773434	0.0	1.00	8.0	33.00	199.0
	MntMeatProducts	2240.0	166.950000	225.715373	0.0	16.00	67.0	232.00	1725.0
	MntFishProducts	2240.0	37.525446	54.628979	0.0	3.00	12.0	50.00	259.0
	MntSweetProducts	2240.0	27.062946	41.280498	0.0	1.00	8.0	33.00	263.0

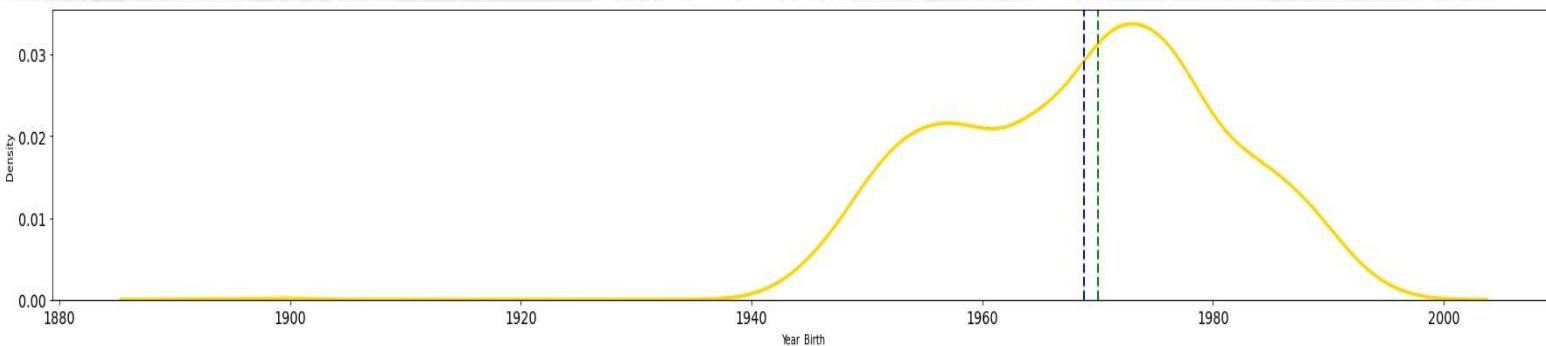
7 Kolom ini tipenya akan diubah menjadi boolean

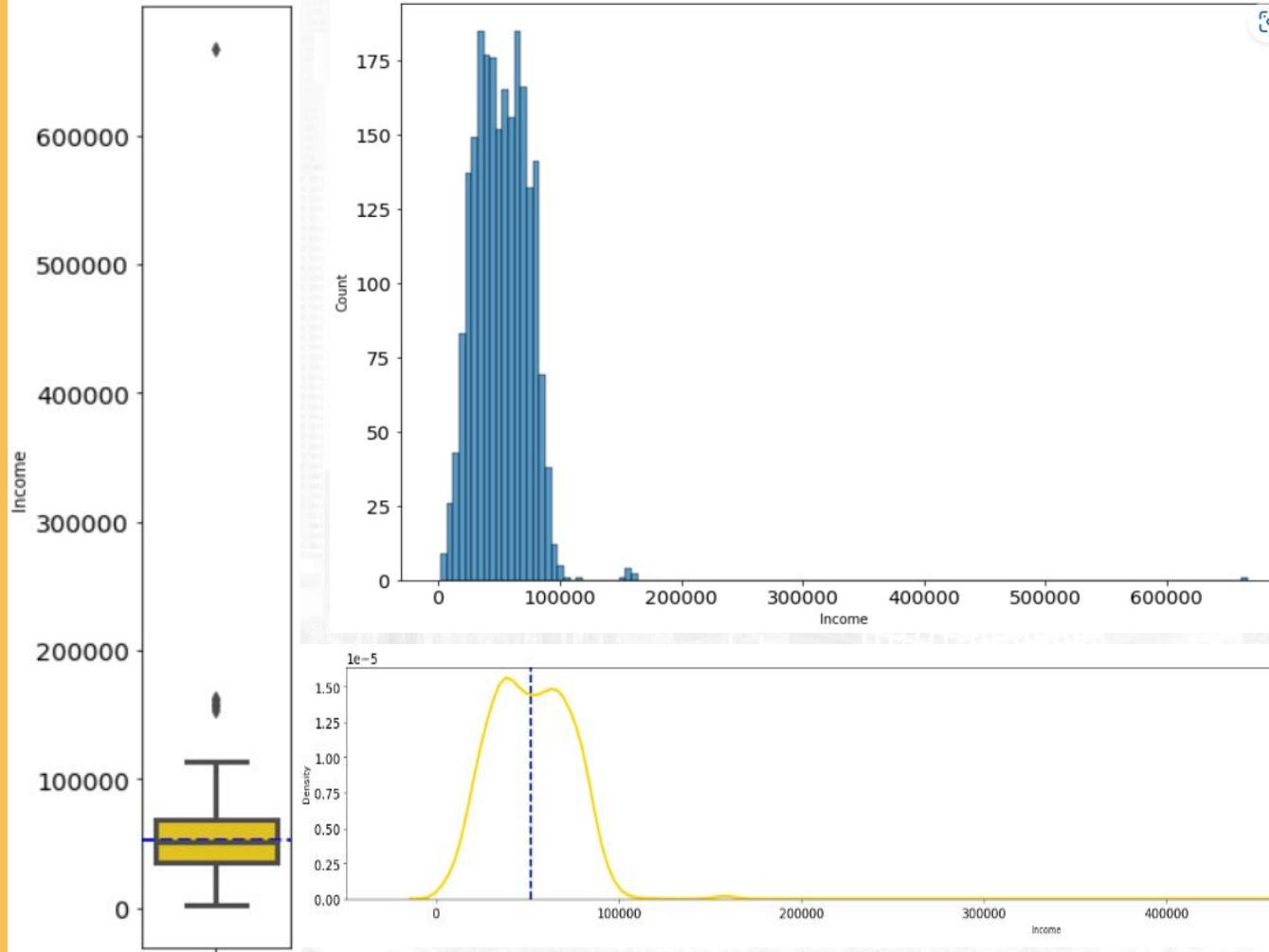
NO 2

Gunakan visualisasi untuk melihat distribusi masing-masing kolom (feature maupun target). Tuliskan hasil observasinya, misalnya jika ada suatu kolom yang distribusinya menarik (misal skewed, bimodal, ada outlier, ada nilai yang mendominasi, kategorinya terlalu banyak, dsb). Jelaskan juga apa yang harus di-follow up saat data pre-processing.

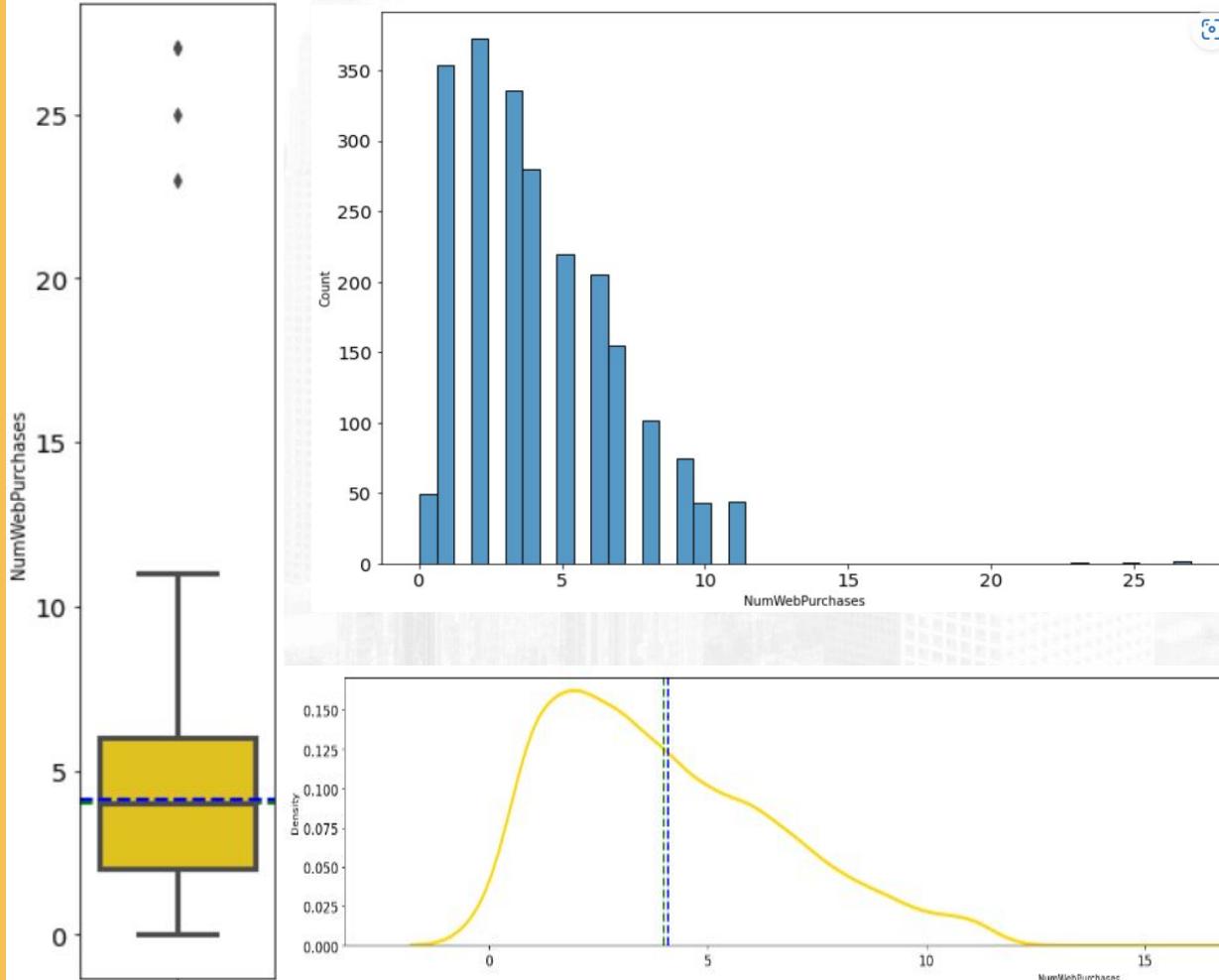


- Kolom '**year_birth**' terdistribusi negative skew karena mean lebih kecil dibandingkan nilai mediannya (ada user yang tahun kelahirannya sangat jauh dibanding user lainnya). Untuk kolom tersebut bisa diproses agar terdistribusi normal



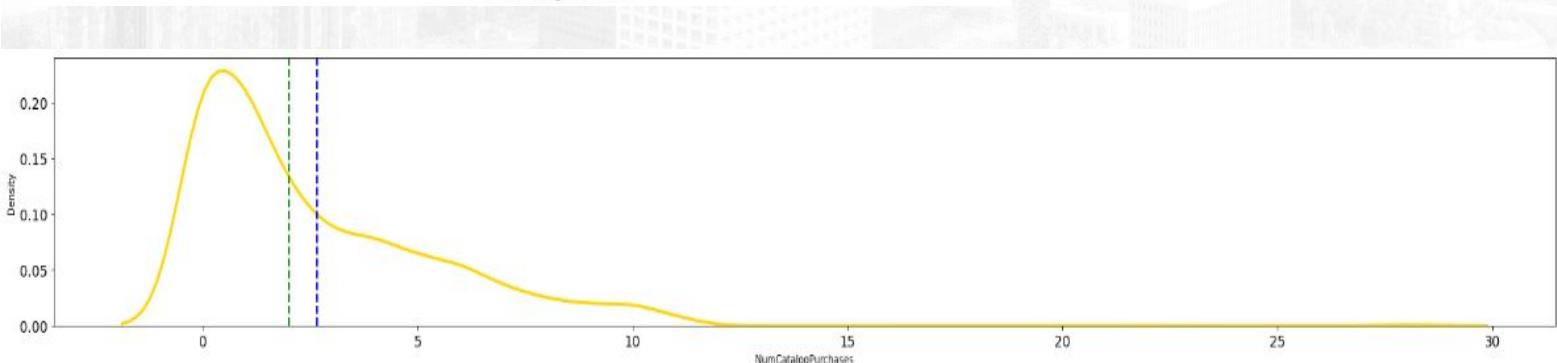
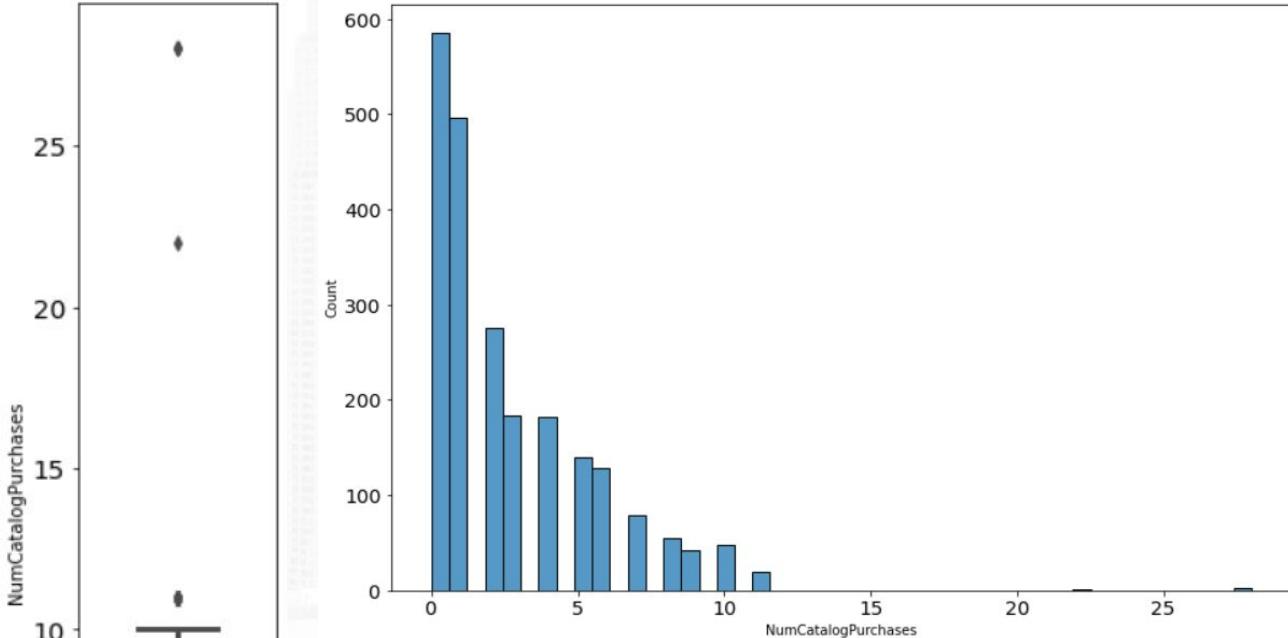


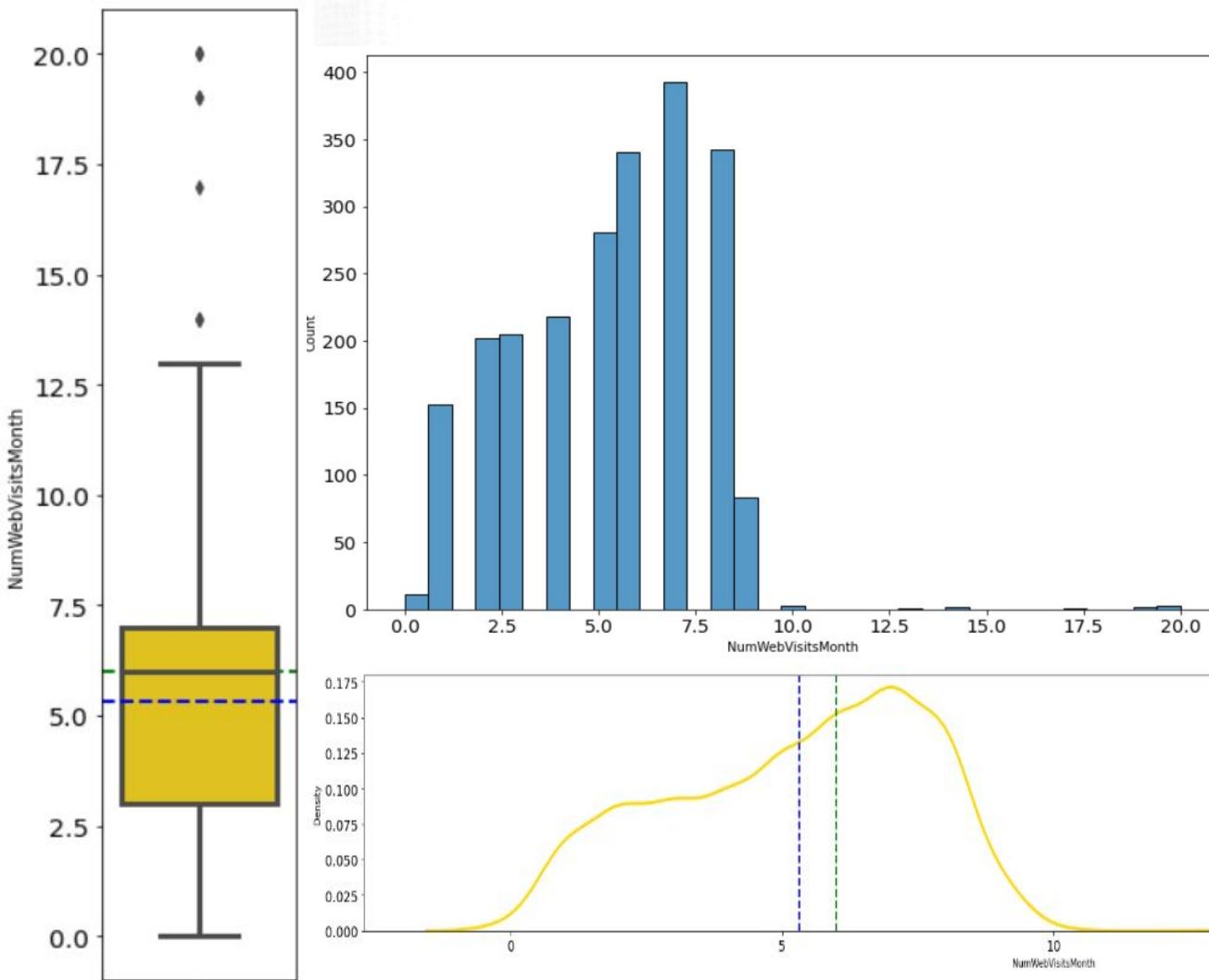
- Kolom ‘income’ memiliki outlier karena nilai maxnya sangat jauh dibandingkan dari nilai mean & median. Nilai outlier kolom ini perlu diproses pada tahap selanjutnya



Kolom
'NumWebPurchases'
 terdapat outliers, mean
 hampir sama dengan
 median dan terdistribusi
 normal

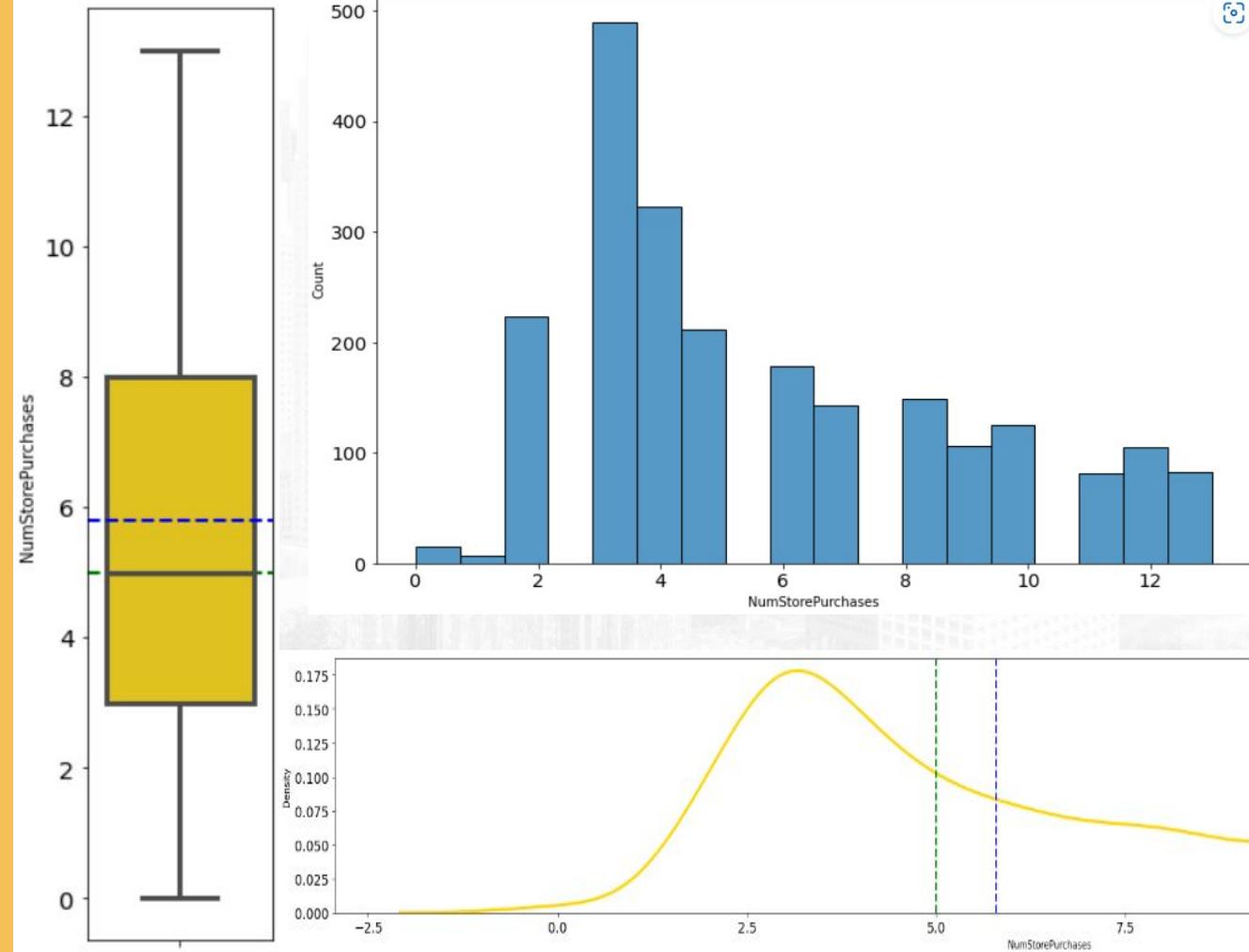
Kolom
'NumCatalogPurchase
s', terdapat outliers,
mean > median,
terdistribusi positif
skew

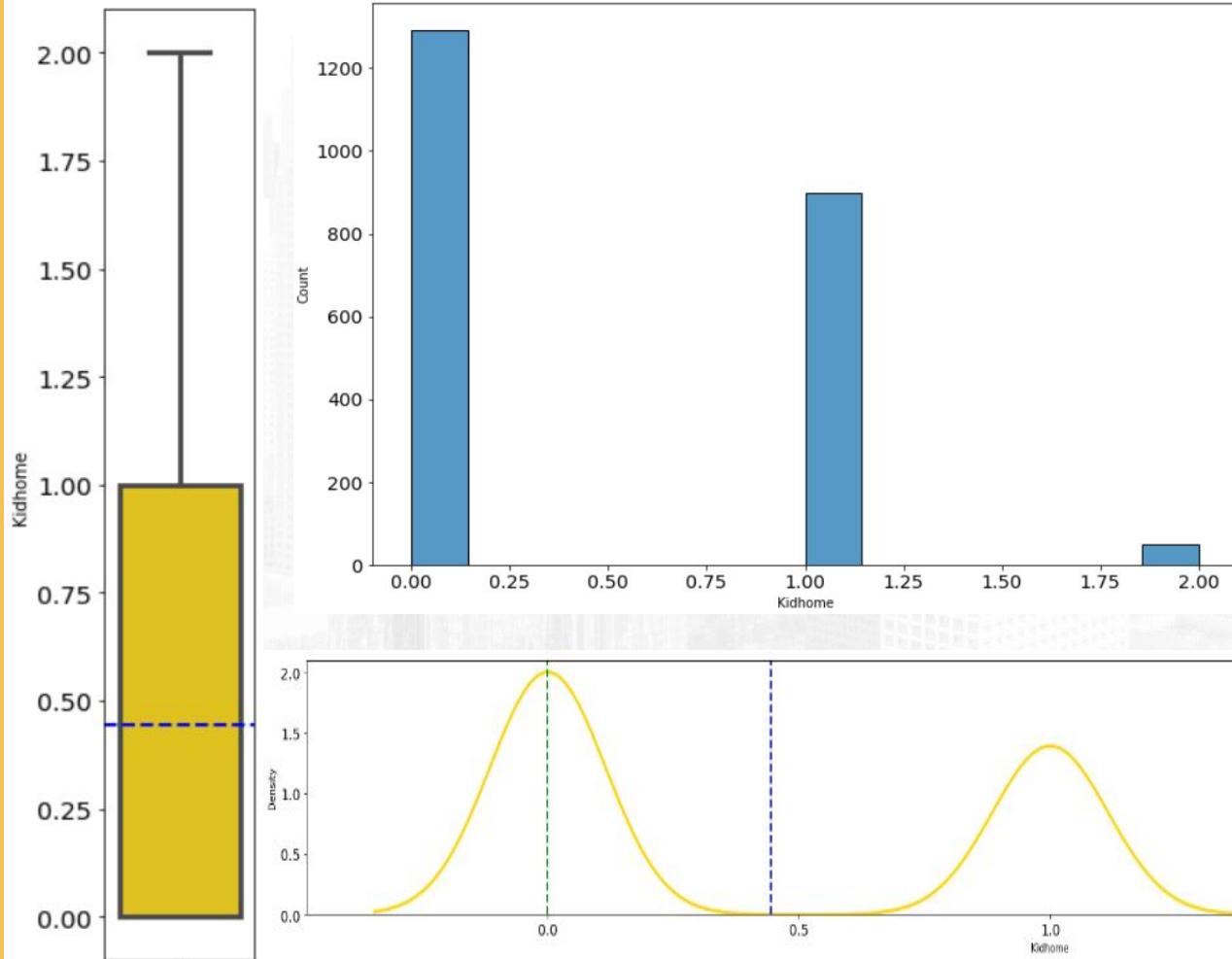




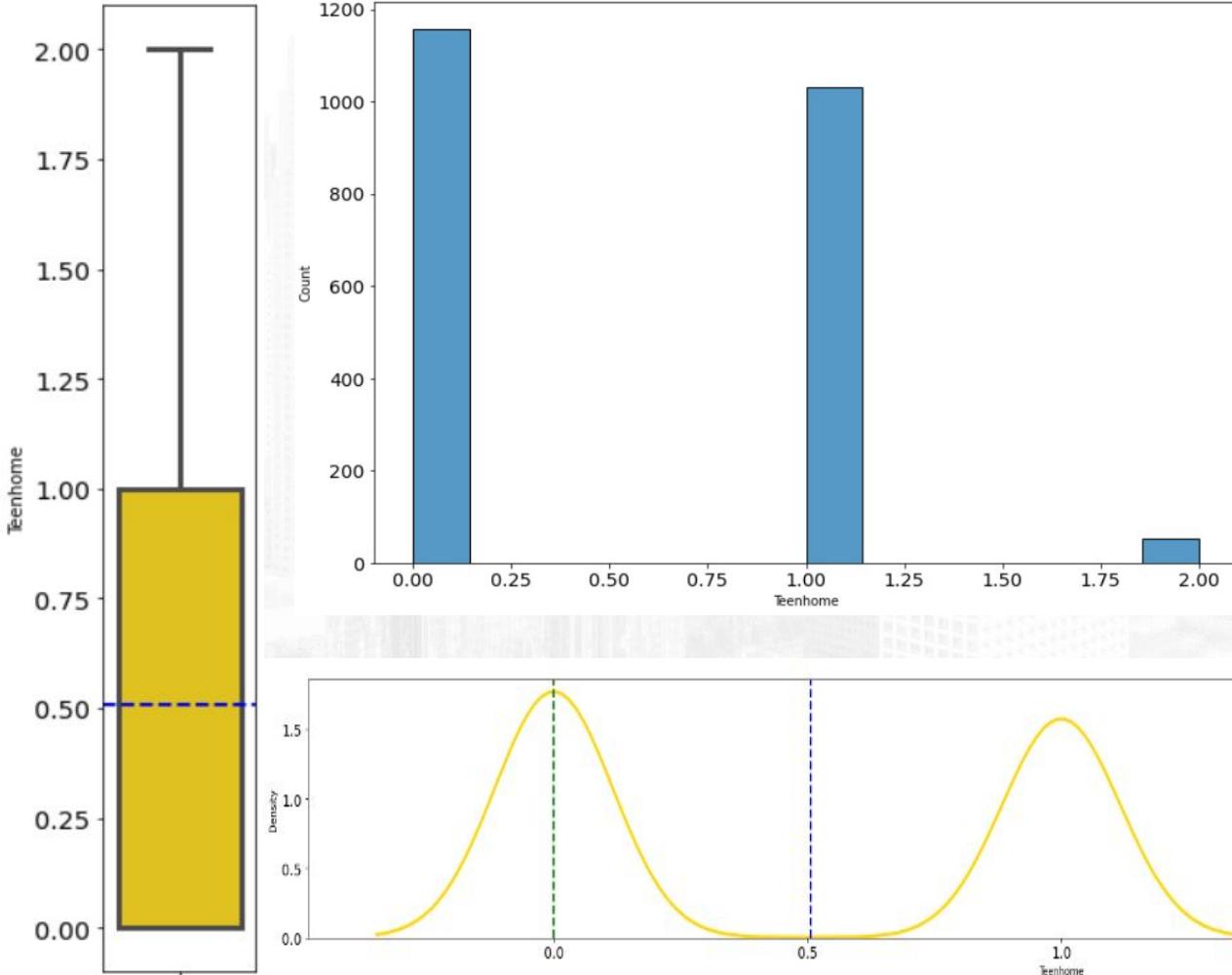
Kolom
'NumWebVisitsMonth'
terdapat outliers, mean <
median terdistribusi negatif
skew

Kolom
'NumStorePurchases'
 tidak ada outliers, mean > median, terdistribusi positif skew

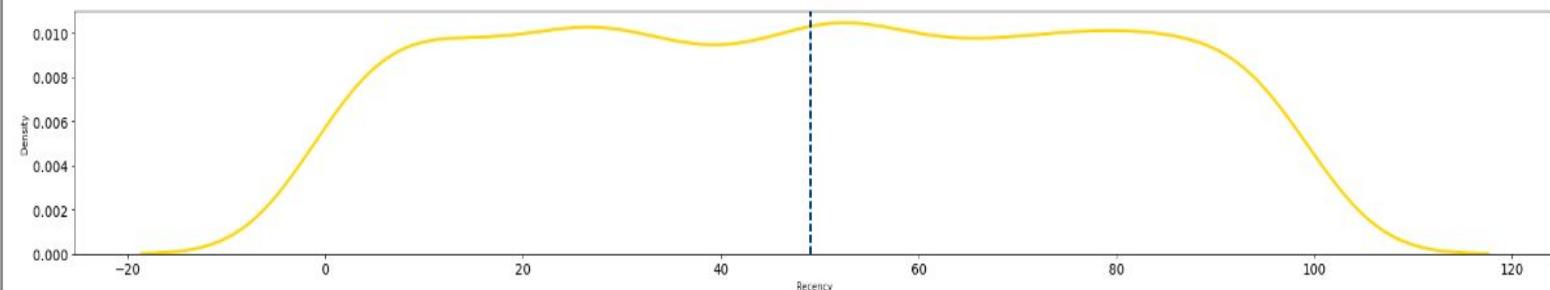
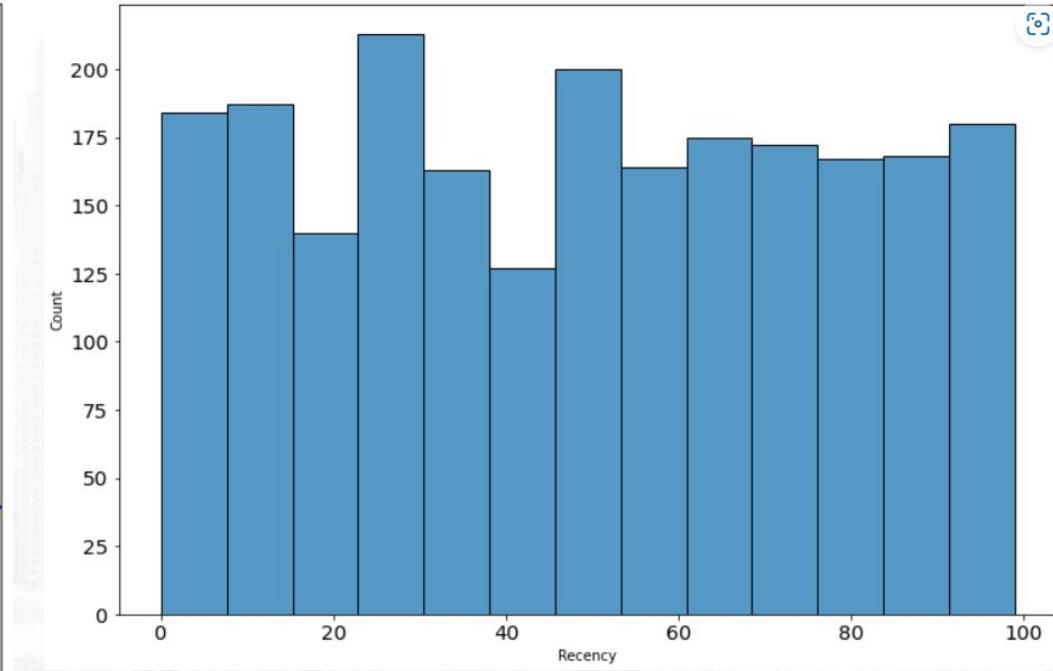
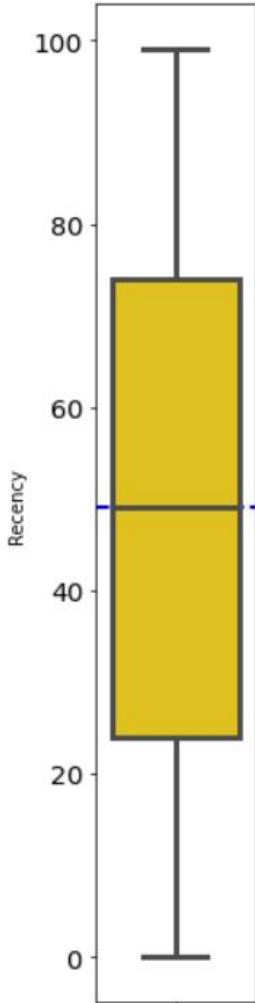




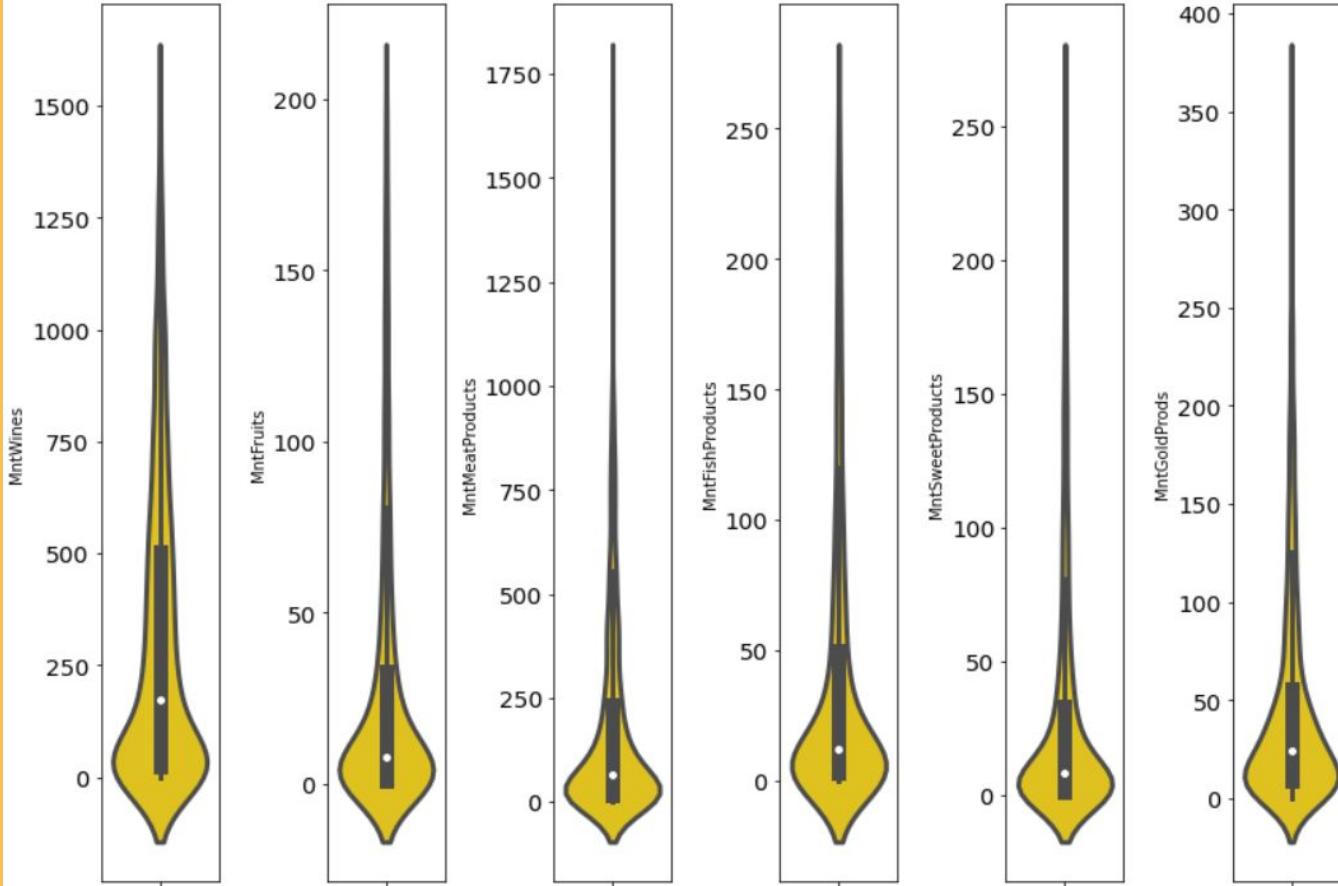
Kolom '**kidhome**' dan '**teenhome**' terdistribusi bimodal karena ada 2 puncak (nilai modus tertinggi) pada user yang memiliki 1 anak dan tidak punya anak



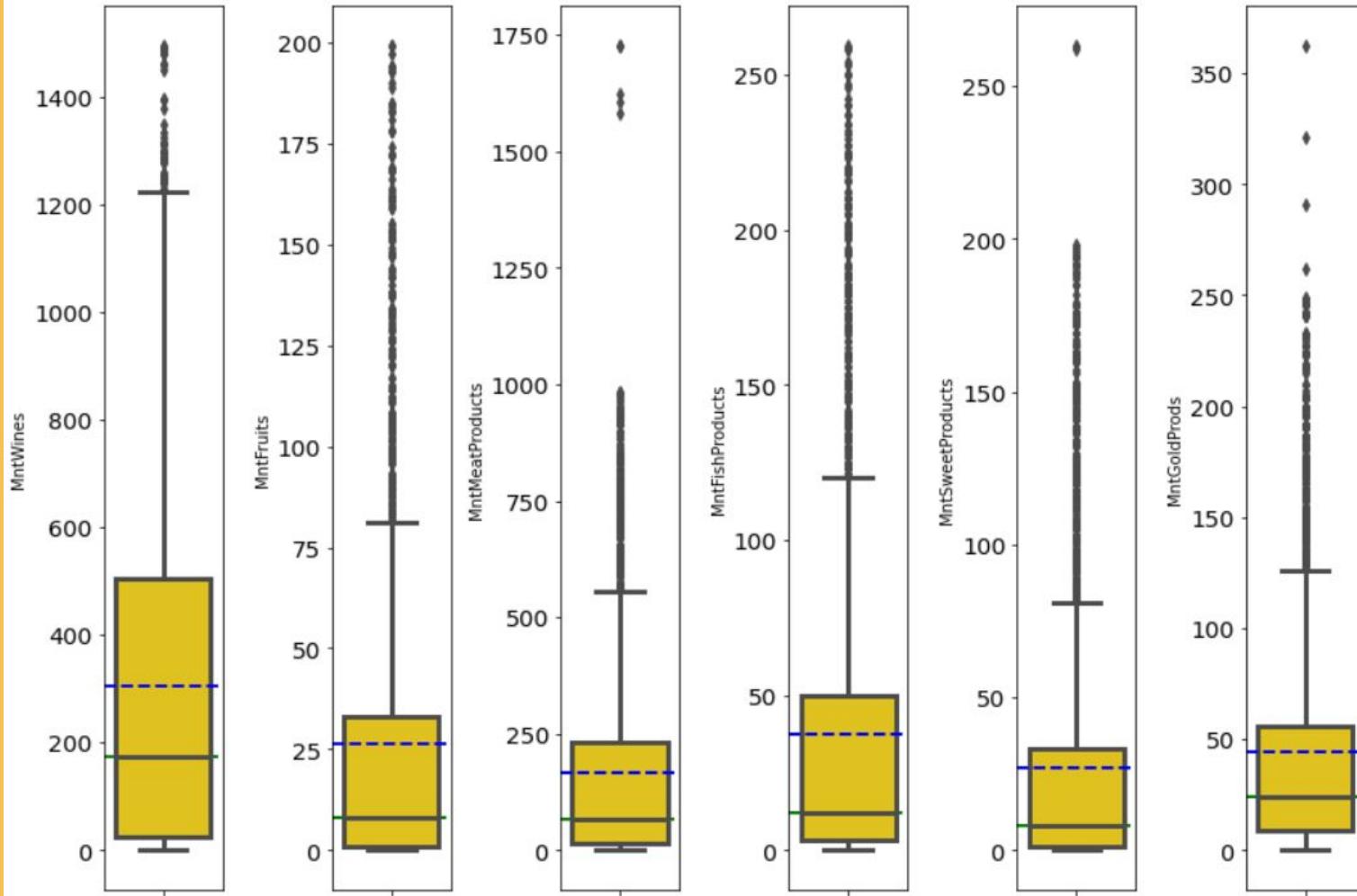
Kolom '**kidhome**' dan '**teenhome**' terdistribusi bimodal karena ada 2 puncak (nilai modus tertinggi) pada user yang memiliki 1 anak dan tidak punya anak



Kolom 'recency'
terdistribusi normal

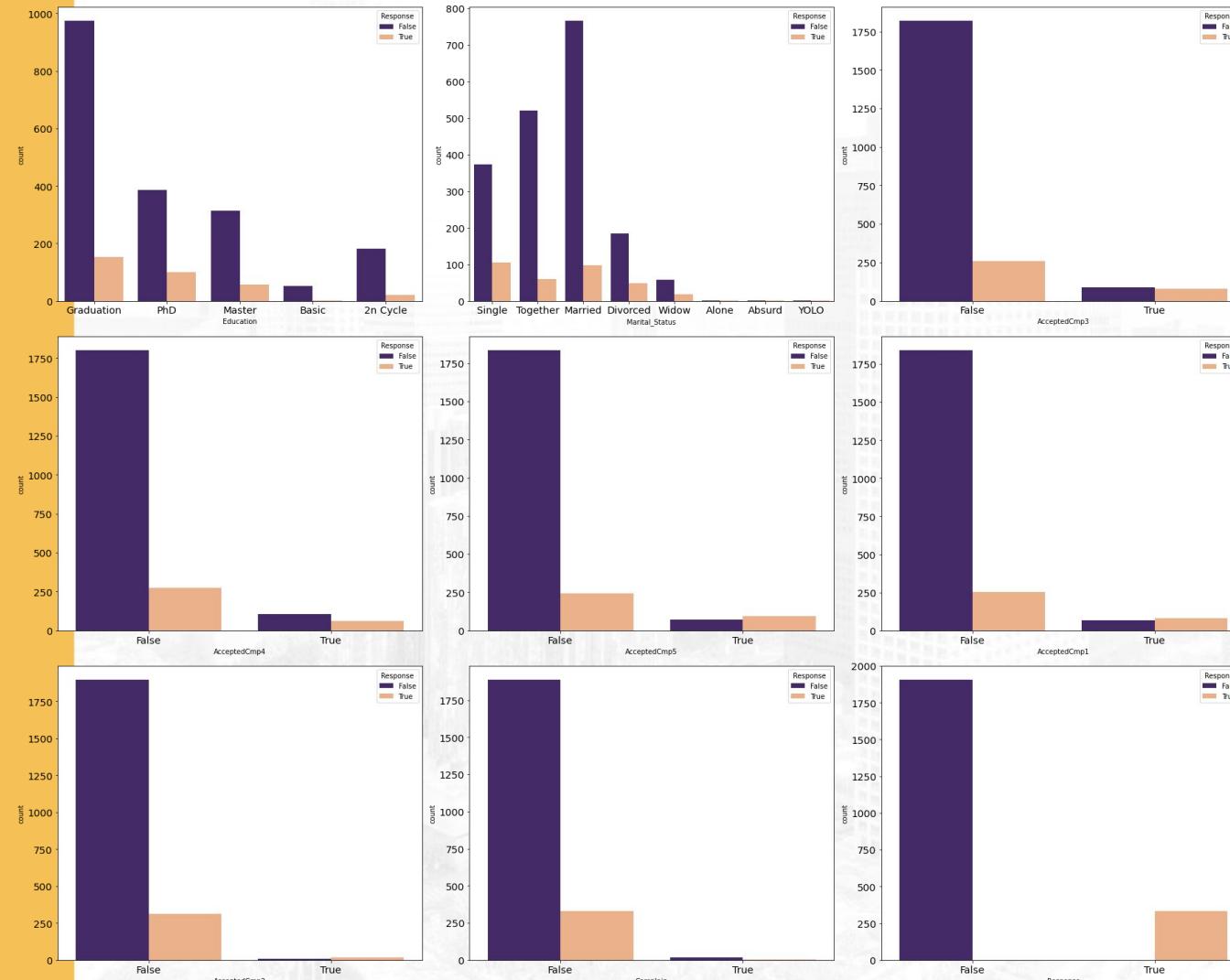


Seluruh kolom 'amount products' terdistribusi positive skew karena nilai mean lebih besar dibanding nilai mediannya, namun terdapat outlier yang perlu diproses pada tahap selanjutnya



Keseluruhan outliers produk ini akan diproses pada tahap selanjutnya.

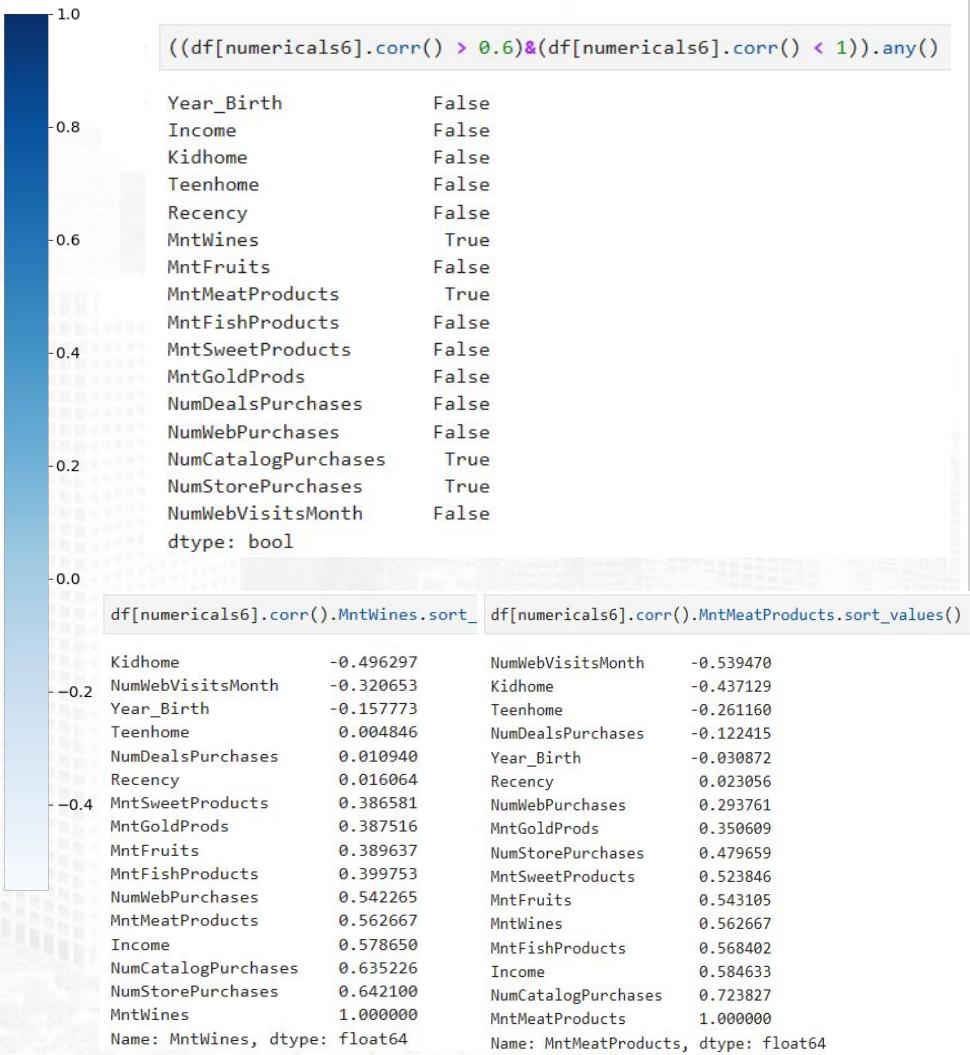
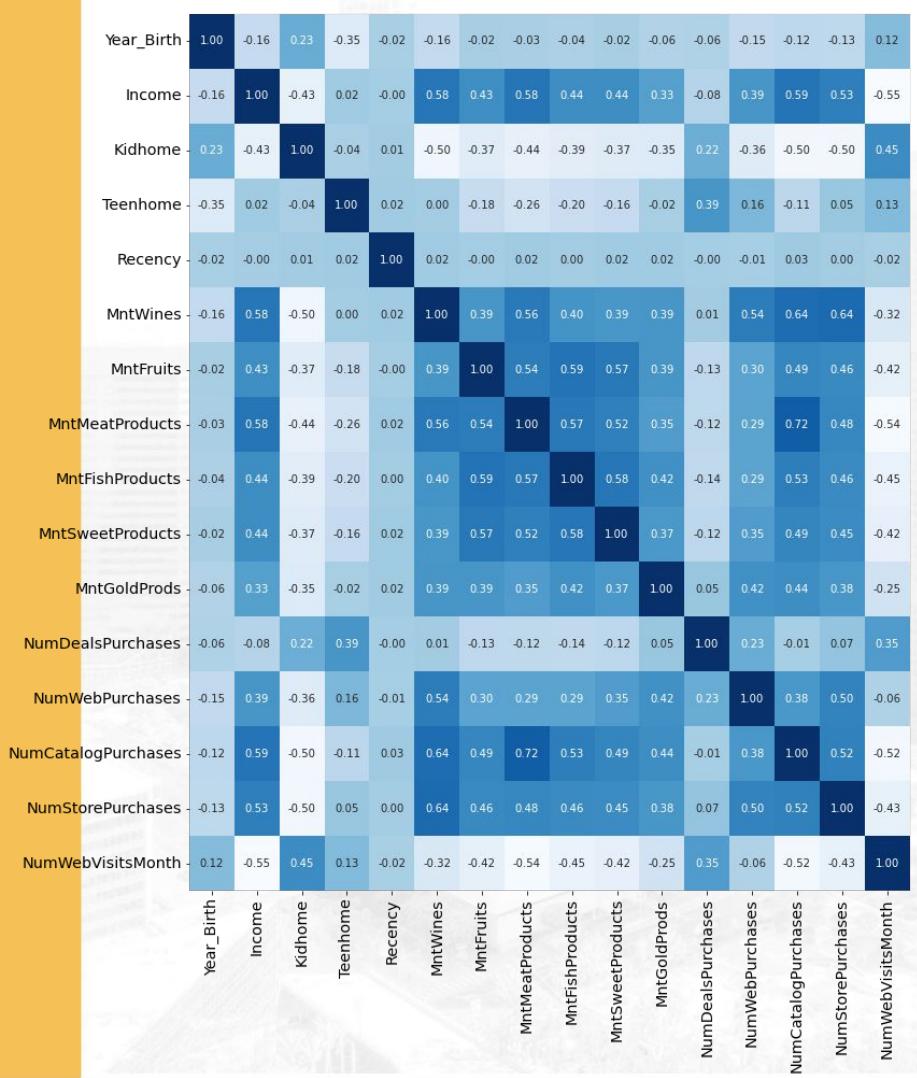
Kolom lainnya terdistribusi positive skew karena mean > median

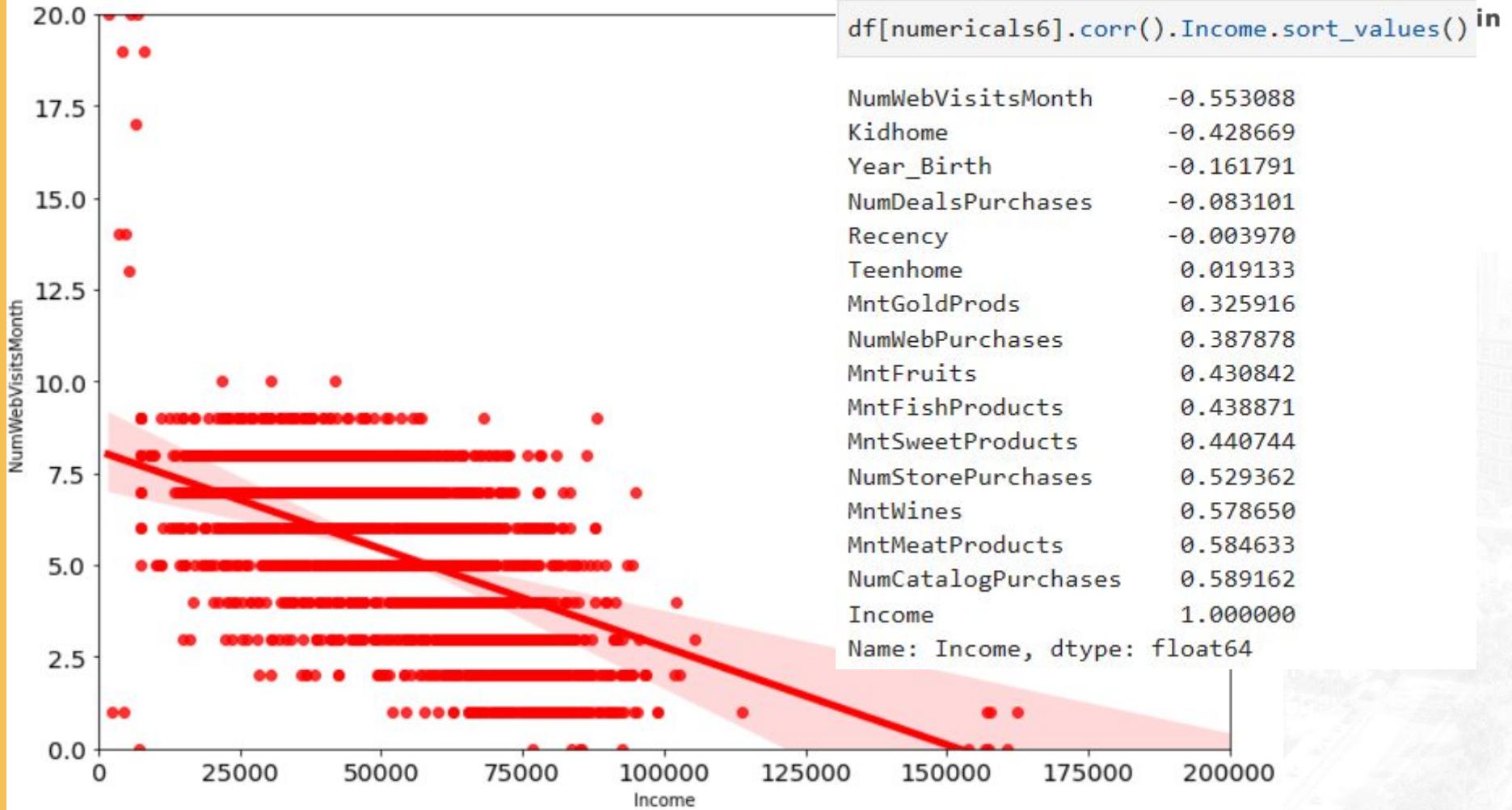


NO 3

A. Bagaimana korelasi antara masing-masing feature dan label. Kira-kira feature mana saja yang paling relevan dan harus dipertahankan?

11 feature yang dipertahankan: '**Income**', '**MntWines**', '**MntFruits**', '**MntMeatProducts**', '**MntFishProducts**', '**MntSweetProducts**', '**MntGoldProds**', '**NumWebPurchases**', '**NumCatalogPurchases**', '**NumStorePurchases**', '**Marital_Status**', '**Education**'





B. Bagaimana korelasi antar-feature, apakah ada pola yang menarik? Apa yang perlu dilakukan terhadap feature itu?

Korelasi antar feature:

- Korelasi antara kolom '**Income**' dan seluruh kolom 'amount products' berada pada nilai korelasi di atas 0.44, kecuali kolom '**MntGoldProds**', sehingga terlihat trend bahwa semakin tinggi income maka semakin tinggi pembelian produk
- Korelasi antara kolom '**Income**' dan keseluruhan kolom purchase ('**NumWebPurchases**', '**NumCatalogPurchases**', '**NumStorePurchases**'), semuanya berada di atas 0.53 kecuali pada kolom '**NumWebPurchases**' yang berarti pelanggan dengan income tinggi lebih sering melakukan pembelian melalui toko dan katalog
- Kolom '**Income**' tidak berkorelasi dengan kolom '**NumDealsPurchases**' dan '**NumWebVisitsMonth**'

NO 4

Selain EDA, lakukan juga beberapa analisis dan visualisasi untuk menemukan suatu business insight. Tuliskan minimal 3 insight, dan berdasarkan insight tersebut jelaskan rekomendasinya untuk bisnis.

Business insight

- Semakin tinggi pendapatan pelanggan maka semakin banyak jumlah produk yang dibeli
- Pelanggan dengan pendapatan tinggi lebih suka berbelanja melalui store dan katalog
- Orang yang menerima campaign terakhir (accepted campaign 5) banyak berasal dari golongan income tinggi
- Orang dengan income tinggi paling banyak melakukan pembelian melalui katalog dengan produk favorit wine dan meat
- Pembeli terbanyak berada dalam kategori Married,Together,dan single

Business recommendation

Next campaign bisa ditargetkan ke orang dengan income yang tinggi dan difokuskan pada pembelian melalui katalog berkategori wine dan meat

NO 5

A. Buat Repository Git

<https://github.com/marhanura/finprokel2/tree/stage1>

B. Upload file notebook atau file penggerjaan lainnya pada repository tersebut

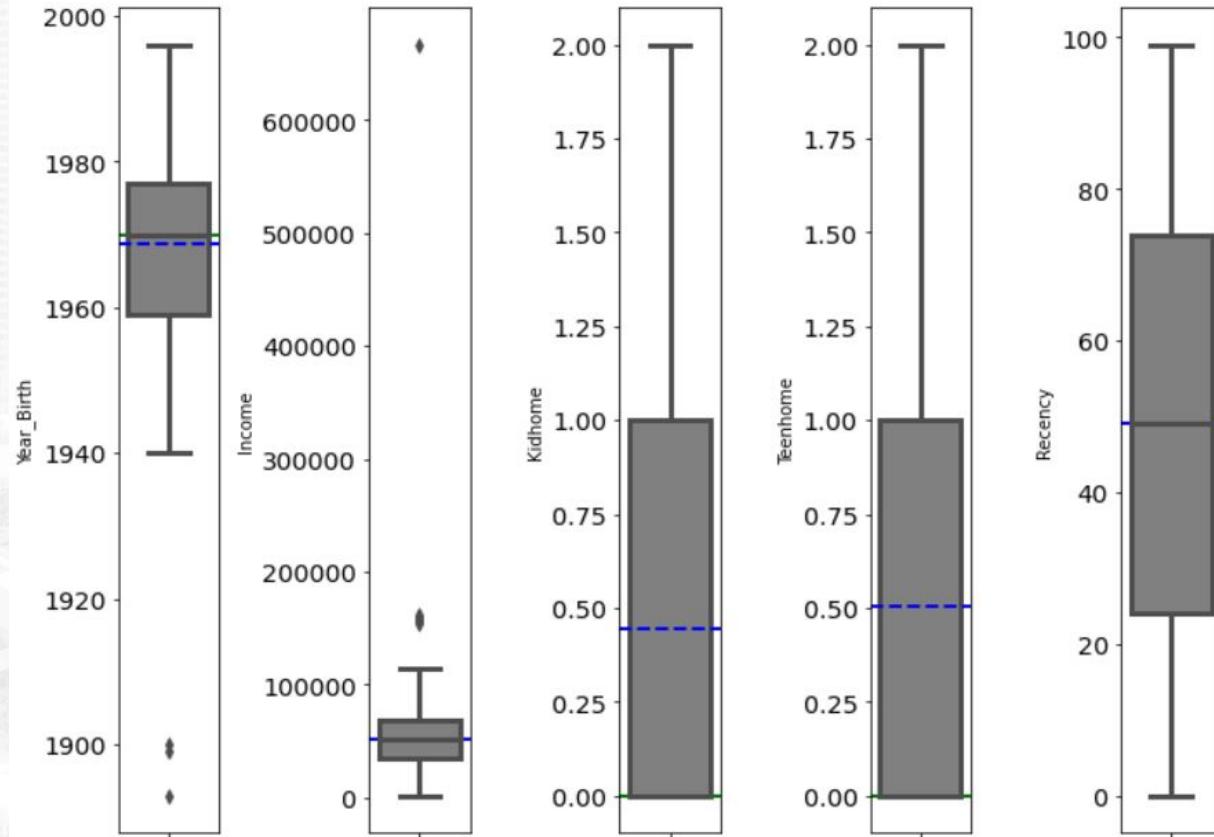
Repository Git:

<https://github.com/marhanura/finprokel2/tree/stage1>

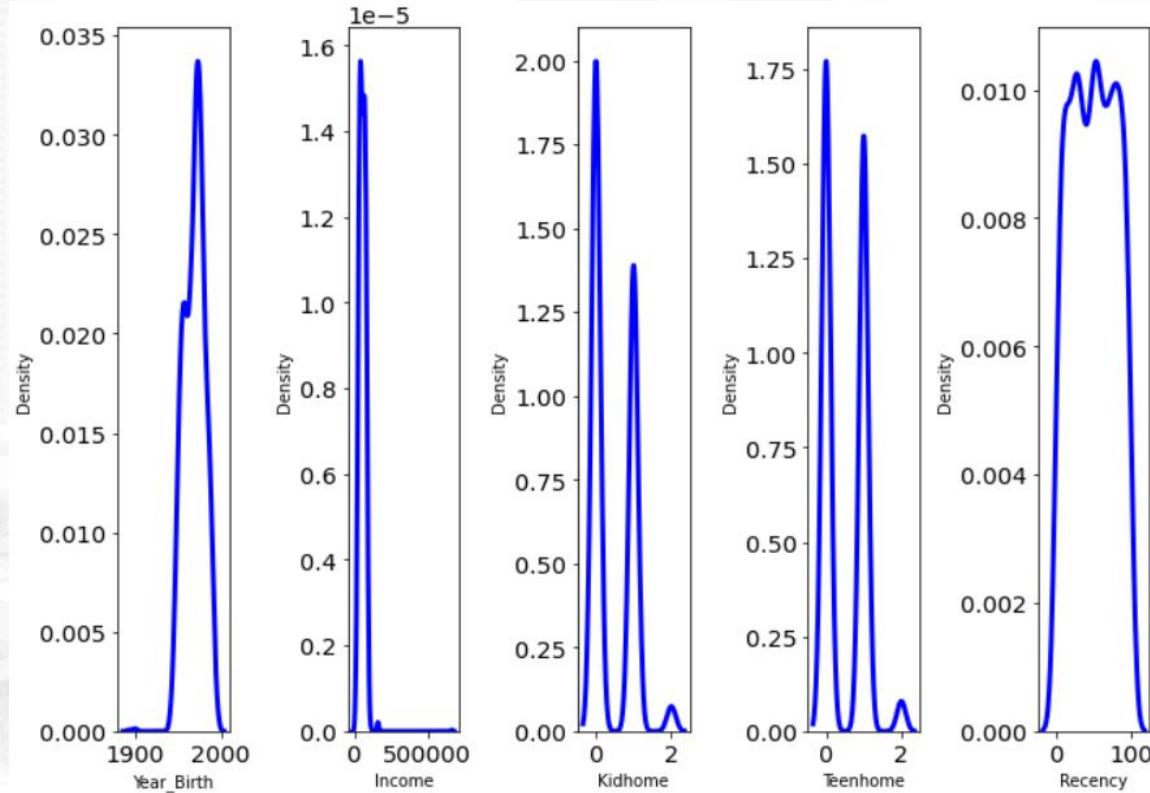
Link Google Drive Stage 1

<https://drive.google.com/drive/u/2/folders/1bto6vFxH-eqbT0gw45fe8laSwrUjNfG8>

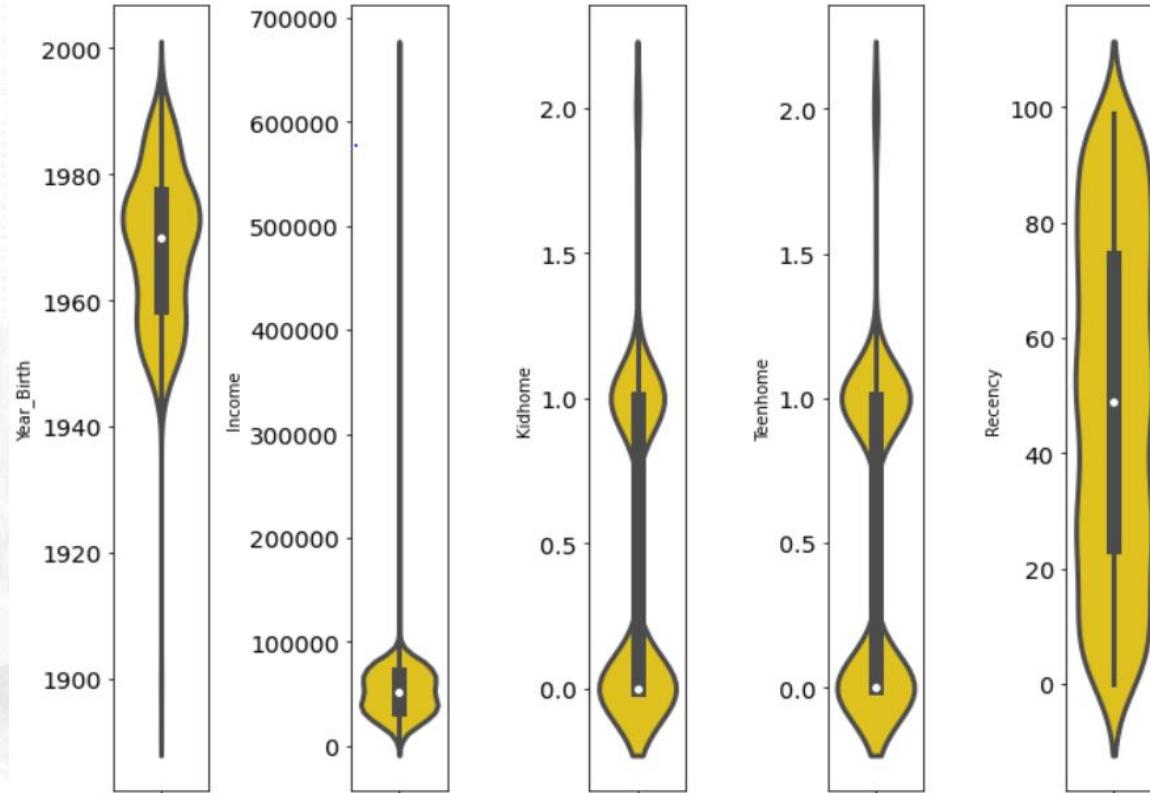
Visualisasi Tambahan



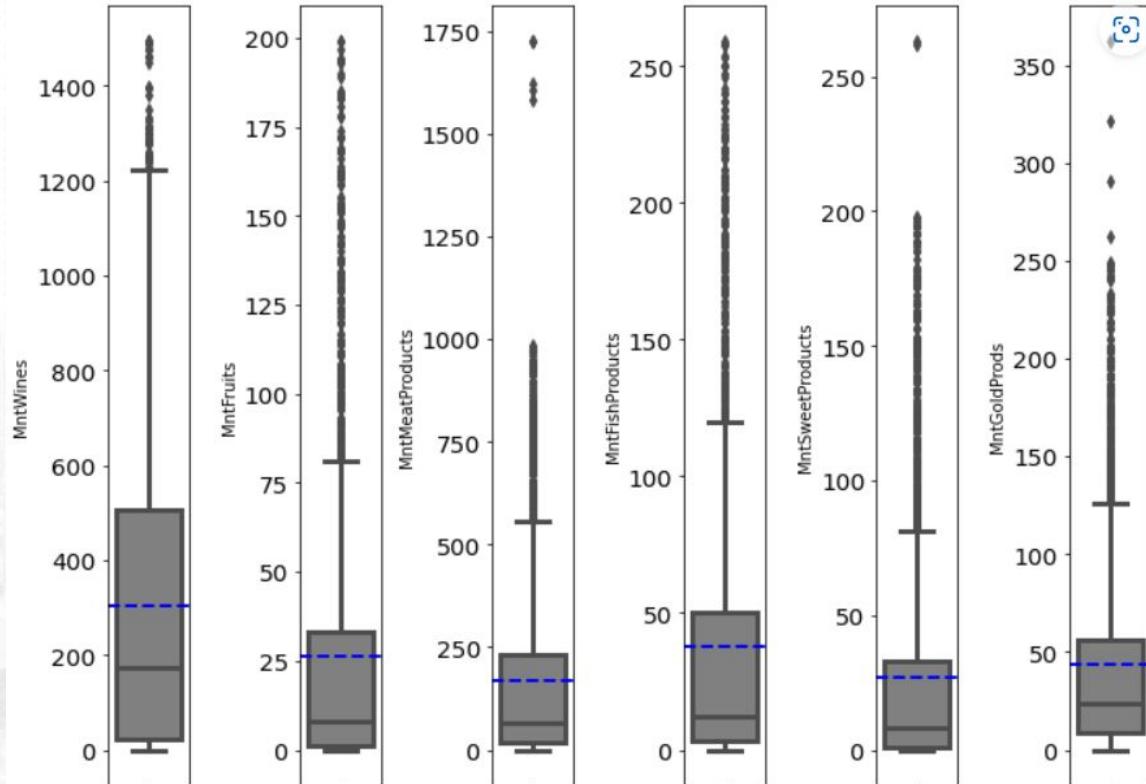
Visualisasi Tambahan



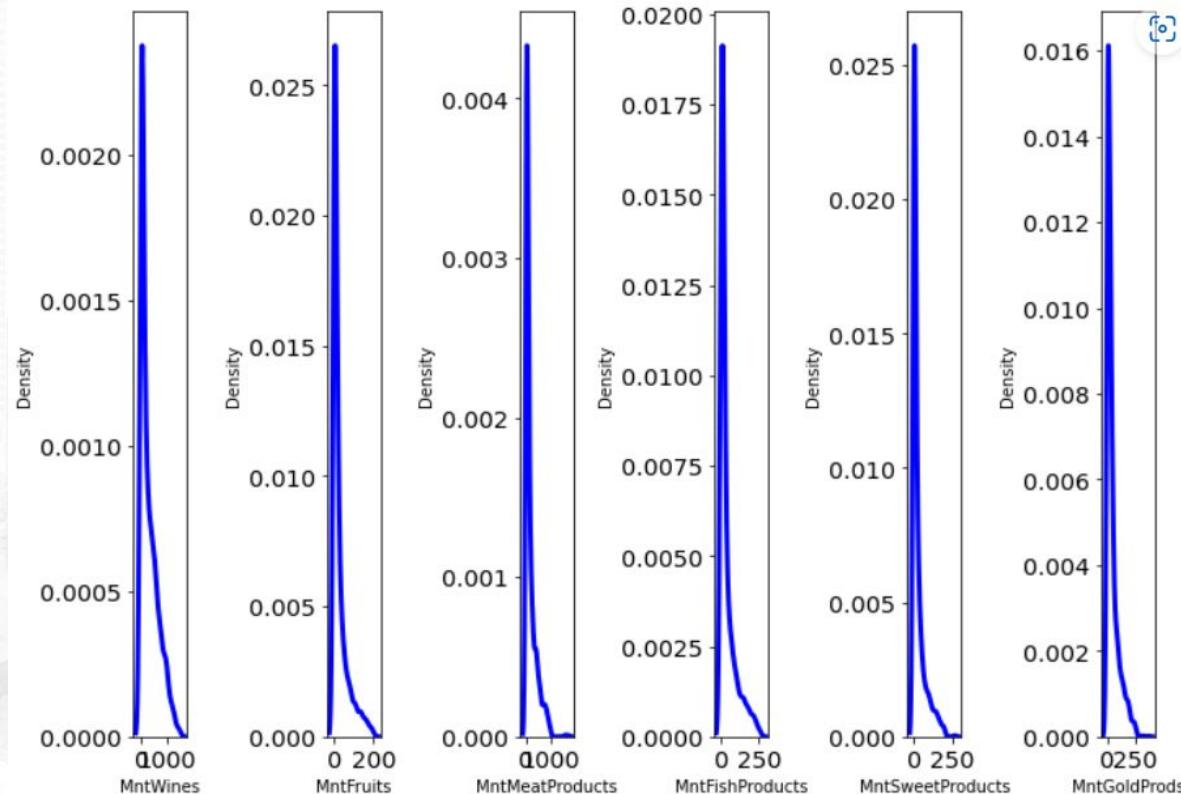
Visualisasi Tambahan



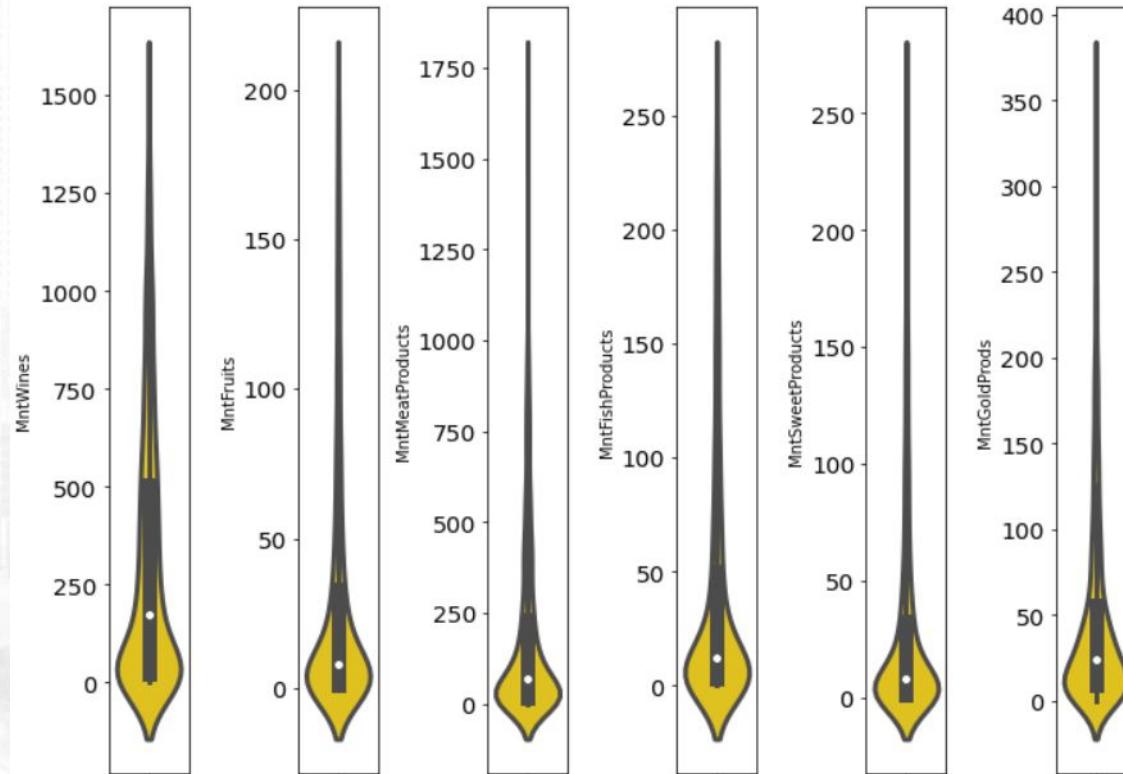
Visualisasi Tambahan



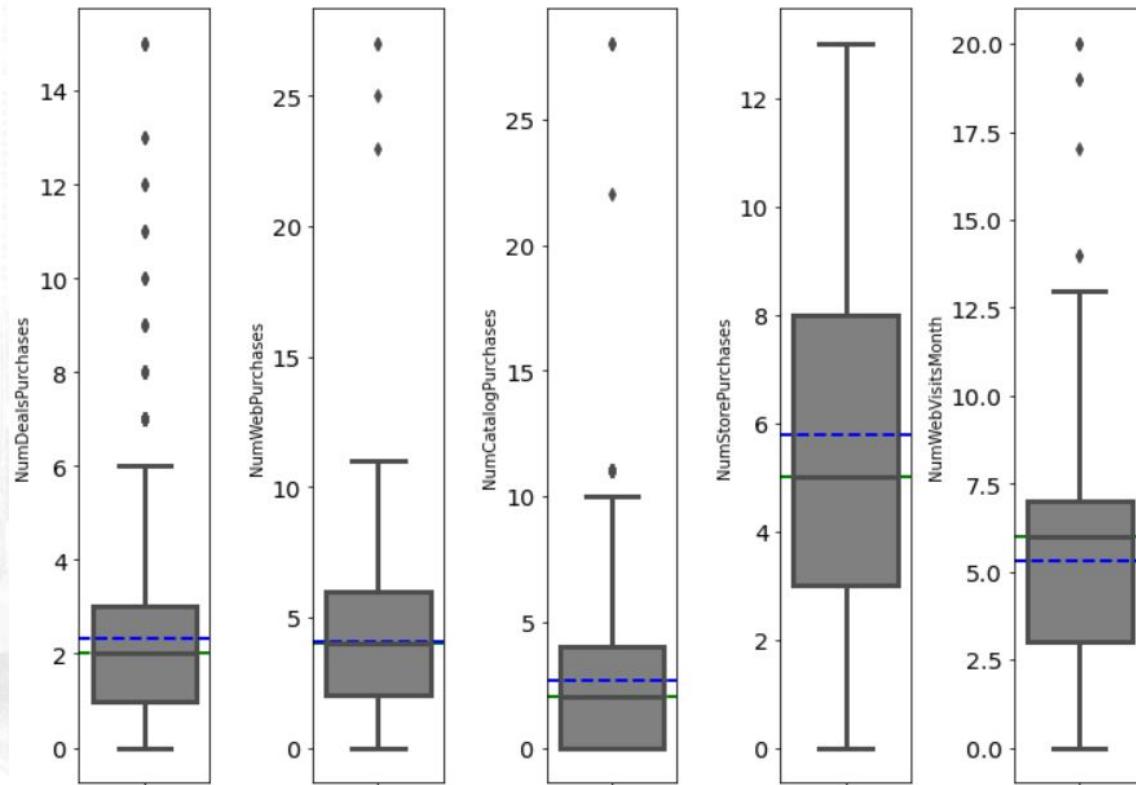
Visualisasi Tambahan



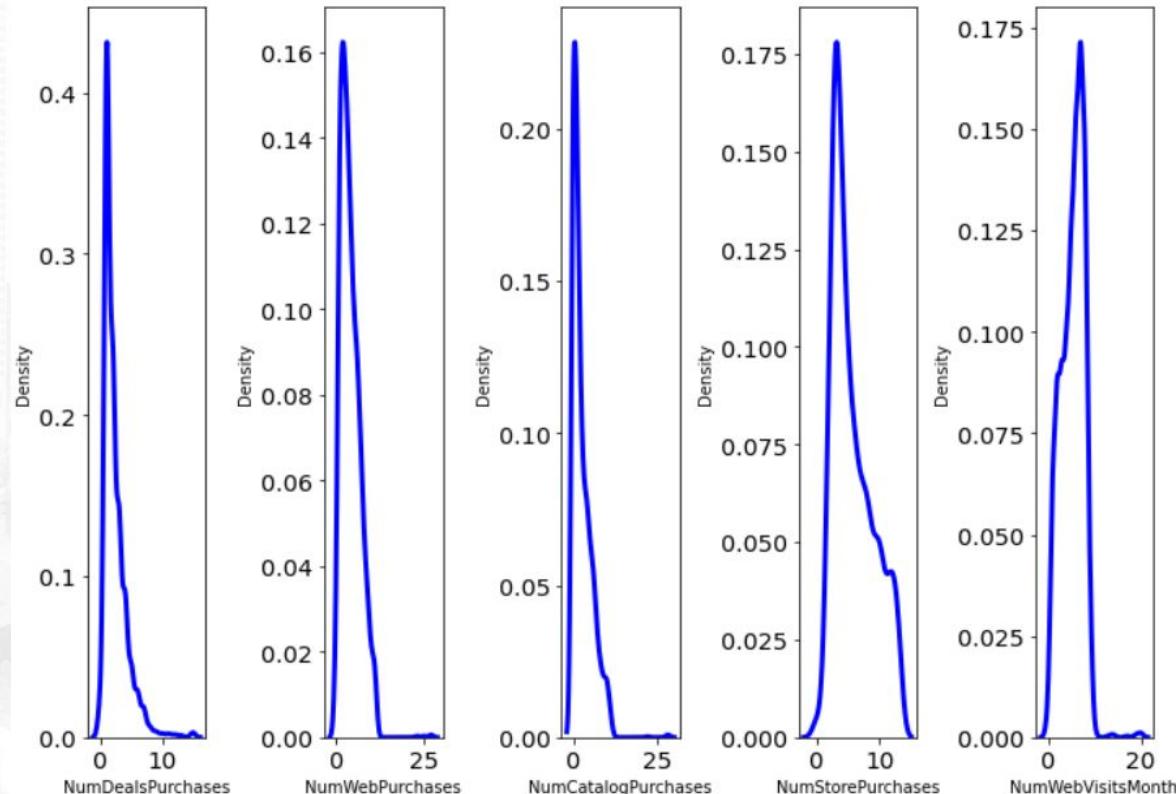
Visualisasi Tambahan



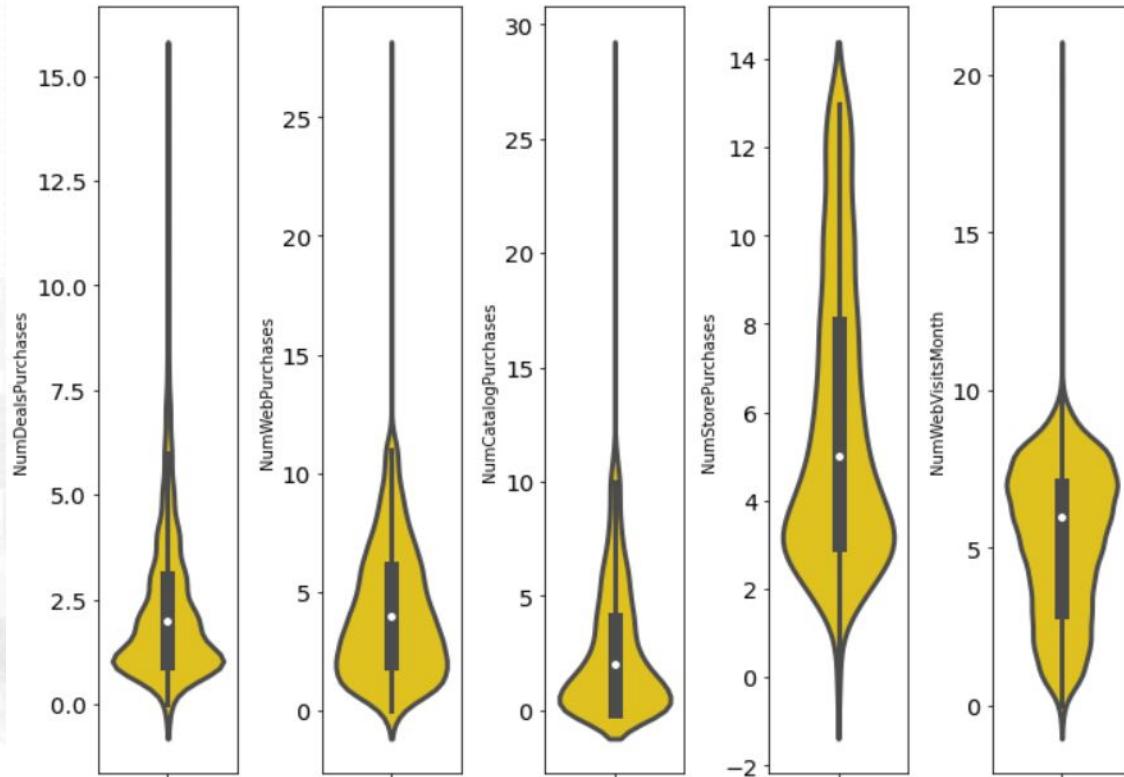
Visualisasi Tambahan



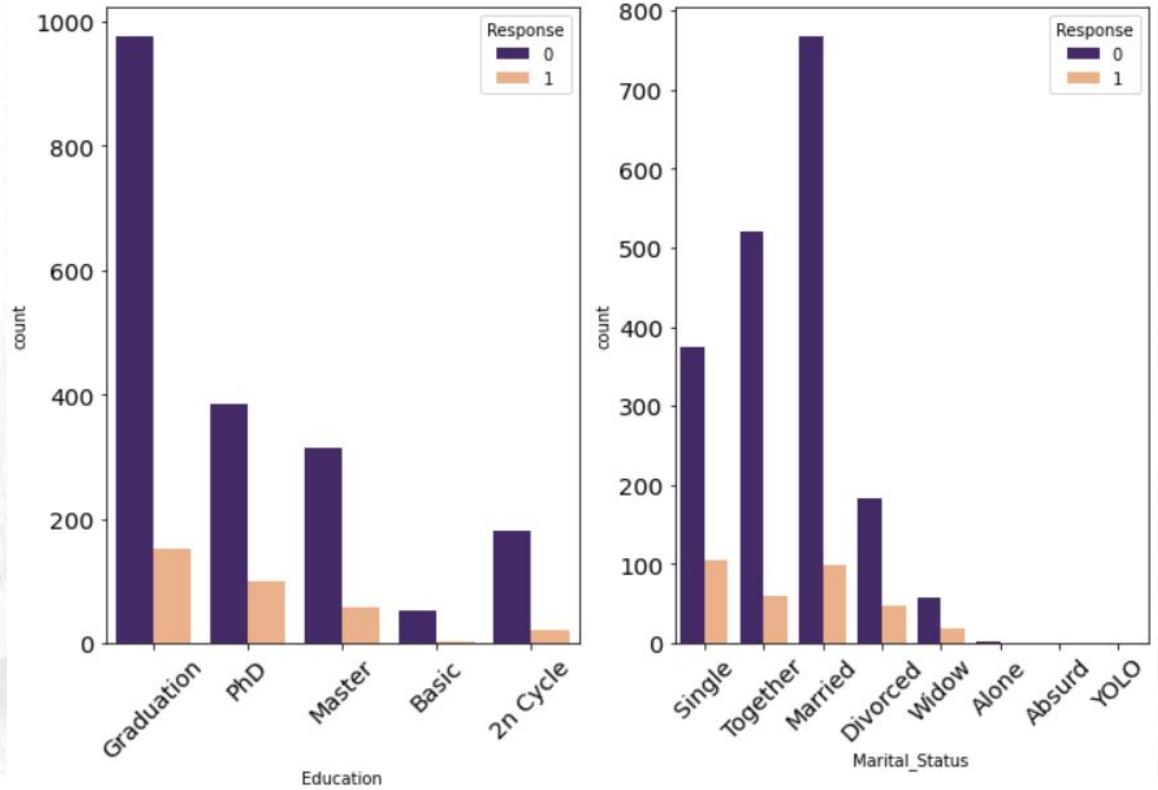
Visualisasi Tambahan



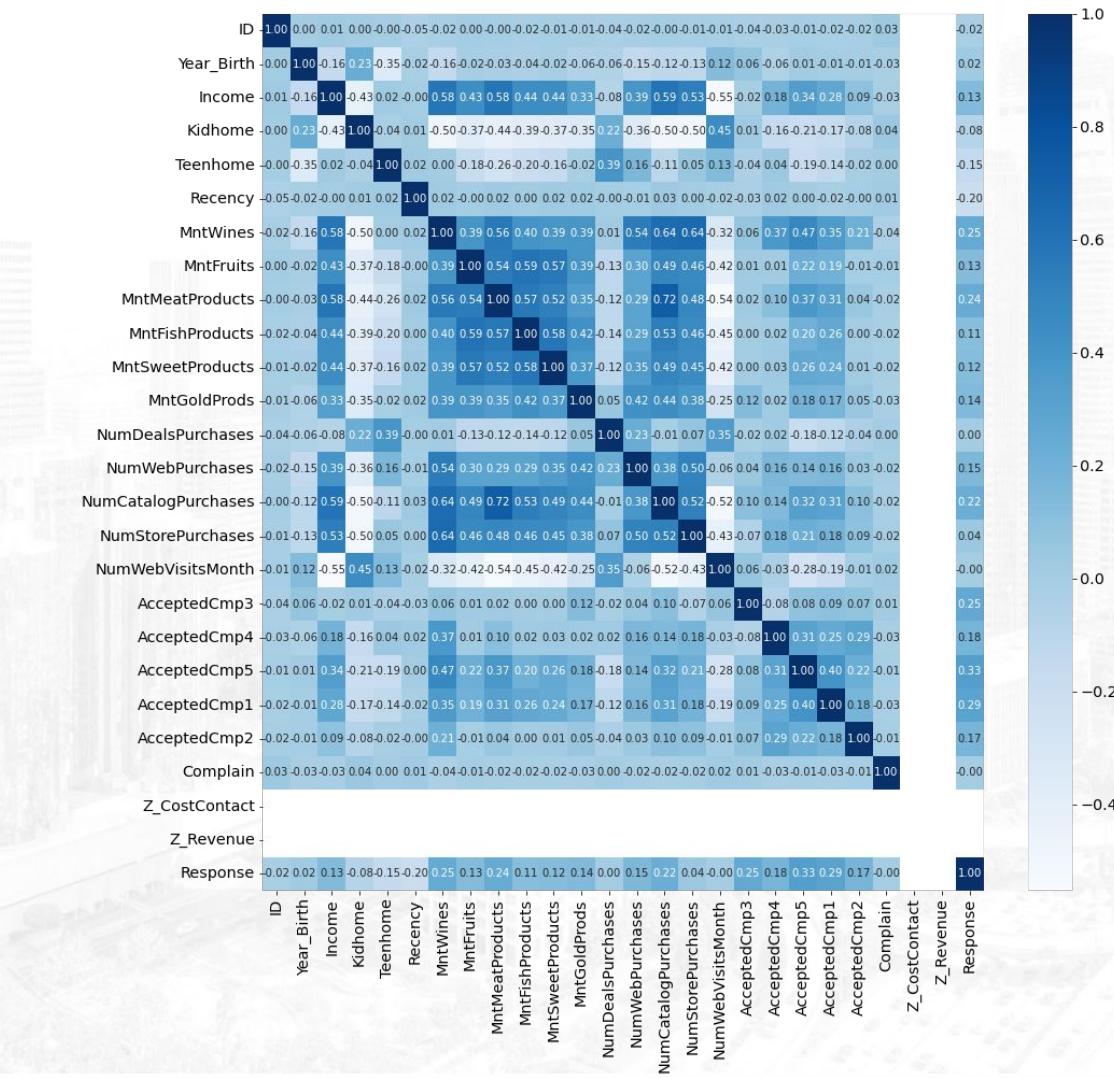
Visualisasi Tambahan



Visualisasi Tambahan



Visualisasi Tambahan



Visualisasi Tambahan

```
((df.corr()>0.7)&(df.corr()<1)).any()
```

ID	False
Year_Birth	False
Income	False
Kidhome	False
Teenhome	False
Recency	False
MntWines	False
MntFruits	False
MntMeatProducts	True
MntFishProducts	False
MntSweetProducts	False
MntGoldProds	False
NumDealsPurchases	False
NumWebPurchases	False
NumCatalogPurchases	True
NumStorePurchases	False
NumWebVisitsMonth	False
AcceptedCmp3	False
AcceptedCmp4	False
AcceptedCmp5	False
AcceptedCmp1	False
AcceptedCmp2	False
Complain	False
Z_CostContact	False
Z_Revenue	False
Response	False

Jumlah pembelian ‘daging’ tidak sama dengan jumlah pembayaran melalui catalog, sehingga kedua kolom ini bukanlah kolom redundan

