



# Analisis Supermarket Customers

Dewi Maharani Dyah Pitaloka



# Lingkup Pembahasan

Latar Belakang

Tujuan Analisis

Masalah Yang Akan  
Diselesaikan

Aksi Yang Diharapkan

Data Understanding

Pengolahan Data

- Cleaning Data
- Exploratory Data Analysis

Kesimpulan

Rekomendasi



# Latar Belakang

Supermarket Purwadhika, sebagai pemain utama di industri ritel Indonesia, telah memantapkan diri sebagai pusat belanja serbaguna, menawarkan produk mulai dari bahan makanan sehari-hari hingga barang mewah. Keanekaragaman produk ini menjadikan Purwadhika tujuan belanja yang disukai oleh berbagai lapisan masyarakat. Dalam menghadapi persaingan yang semakin ketat di industri ritel, Supermarket Purwadhika berkomitmen untuk inovasi dan pertumbuhan berkelanjutan.

Tim Analis Data Supermarket Purwadhika telah menggali data dua tahun terakhir untuk membuka wawasan baru yang akan memperkuat posisi Tim Pemasaran dan Penjualan dalam lingkungan pasar yang dinamis. Analisis ini bertujuan untuk memahami interaksi berbagai kelompok demografi dengan produk dan promosi, membantu tim dalam merancang strategi pemasaran yang lebih personal dan menyesuaikan penawaran produk untuk meningkatkan kepuasan pelanggan.

Selanjutnya, evaluasi terhadap respons pelanggan terhadap kampanye pemasaran sebelumnya akan dijadikan dasar untuk mengoptimalkan strategi pemasaran. Melalui analisis ini, kita berharap dapat menentukan pendekatan yang paling resonan dengan pelanggan, serta meningkatkan efektivitas pesan dan saluran komunikasi kita. Temuan ini diharapkan dapat memberikan panduan strategis bagi Supermarket Purwadhika dalam membuat keputusan bisnis yang lebih tepat dan efektif di masa depan.





# Masalah yang akan diselesaikan

Berdasarkan tujuan analisis yang telah disampaikan, dapat diringkas menjadi poin-poin pertanyaan sebagai berikut.

1

## **Analisis Pola Pengeluaran (Spending Patterns Analysis):**

1. Apakah terdapat perbedaan yang signifikan dalam rata-rata pengeluaran pada produk tertentu (seperti anggur, daging, atau buah) antara berbagai kelompok demografi (usia, pendidikan, status pernikahan, pendapatan)?
2. Bagaimana distribusi pengeluaran produk tertentu (misalnya anggur atau daging) di antara kelompok demografi yang berbeda?

2

## **Efektivitas Kampanye (Campaign Effectiveness):**

1. Bagaimana respons pelanggan terhadap kampanye pemasaran yang berbeda berdasarkan kelompok demografi mereka (usia, pendidikan, status pernikahan, pendapatan)?
2. Apakah ada korelasi antara tingkat keberhasilan kampanye pemasaran dengan karakteristik demografi pelanggan?



## Aksi Yang Diharapkan

- Menyesuaikan strategi pemasaran dan penawaran produk berdasarkan hasil analisis untuk menargetkan kelompok demografi yang relevan.
- Mengoptimalkan stok dan promosi produk di toko sesuai dengan preferensi kelompok demografi yang teridentifikasi.
- Menyesuaikan dan meningkatkan kampanye pemasaran masa depan dengan fokus pada kelompok demografi yang paling responsif.
- Menggunakan insight ini untuk menginformasikan pengembangan produk atau layanan baru yang lebih sesuai dengan kebutuhan pelanggan.



# Data Understanding

Dataset "[Supermarket Customers](#)". berisi data terkait pelanggan supermarket, mencakup aspek-aspek berikut:

## People:

- ID Pelanggan
- Tahun Kelahiran
- Tingkat Pendidikan
- Status Perkawinan
- Pendapatan Tahunan Rumah Tangga
- Jumlah Anak di Rumah
- Jumlah Remaja di Rumah
- Tanggal Pendaftaran Pelanggan
- Keterkaitan Pembelian Terakhir
- Keluhan dalam 2 tahun terakhir

## Produk:

- Pengeluaran untuk Anggur, Buah, Produk Daging, Produk Ikan, Produk Manis, dan Produk Emas dalam 2 tahun terakhir

## Promosi:

- Jumlah Pembelian dengan Diskon
- Penerimaan Penawaran di berbagai kampanye promosi

## Tempat:

- Pembelian melalui Website, Katalog, Toko
- Jumlah Kunjungan Web per Bulan



# Data Understanding

- ID: Nomor identifikasi unik pelanggan
- Year\_Birth: Tahun lahir pelanggan.
- Education: Tingkat pendidikan pelanggan.
- Marital\_Status: Status pernikahan pelanggan.
- Income: Pendapatan tahunan pelanggan.
- Kidhome: Jumlah anak di bawah umur di rumah.
- Teenhome: Jumlah remaja di rumah.
- Dt\_Customer: Tanggal pendaftaran pelanggan.
- Recency: Hari sejak pembelian terakhir.
- MntWines: Pengeluaran untuk wine dalam 2 tahun terakhir.
- MntFruits: Pengeluaran untuk buah 2 tahun terakhir.
- MntMeatProducts: Pengeluaran untuk produk daging 2 tahun terakhir.
- MntFishProducts: Pengeluaran untuk ikan 2 tahun terakhir.
- MntSweetProducts: Pengeluaran untuk produk manis 2 tahun terakhir.
- MntGoldProds: Pengeluaran untuk produk emas 2 tahun terakhir.
- NumDealsPurchases: Jumlah pembelian dengan diskon.
- NumWebPurchases: Jumlah pembelian yang dilakukan melalui situs web perusahaan.
- NumCatalogPurchases: Jumlah pembelian melalui katalog.
- NumStorePurchases: Jumlah pembelian langsung di toko.
- NumWebVisitsMonth: Kunjungan web per bulan.
- AcceptedCmp3, AcceptedCmp4, AcceptedCmp5, AcceptedCmp1, AcceptedCmp2: Respon terhadap kampanye promosi.
- Complain: Apakah pelanggan pernah mengajukan keluhan.
- Z\_CostContact, Z\_Revenue: Kolom tambahan.
- Response: Respon pelanggan terhadap kampanye/penawaran terakhir.

## Analisis Pola Pengeluaran

### 1. Demografi Pelanggan:

- Year\_Birth
- Education
- Marital\_Status
- Income

### 2. Pengeluaran Produk:

- MntWines
- MntFruits
- MntMeatProducts
- MntFishProducts
- MntSweetProducts
- MntGoldProds

## Efektivitas Kampanye

### 1. Demografi Pelanggan (sama seperti di atas):

### 2. Data Kampanye:

- AcceptedCmp1
- AcceptedCmp2
- AcceptedCmp3
- AcceptedCmp4
- AcceptedCmp5
- Response





# Pengolahan Data







# Data Cleaning

## Cek Duplikasi Data

*Hasil pemeriksaan menunjukkan bahwa tidak ada data duplikat dalam dataset. Kita bisa melanjutkan ke tahap pembersihan data berikutnya.*

```
Jumlah baris dan kolom di dataset df adalah (2240, 29)
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2240 entries, 0 to 2239
Data columns (total 29 columns):
#   Column                                Non-Null Count  Dtype
---  -
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
dtypes: float64(1), int64(25), object(3)
memory usage: 507.6+ KB
```



# Data Cleaning

## Cek Format Data

*Dari hasil pemeriksaan sebelumnya diketahui format Dt\_Customers masih berupa Object atau String. Perlu dilakukan perubahan format Dt\_Customers menjadi format datetime.*

ID	int64
Year_Birth	int64
Education	object
Marital_Status	object
Income	float64
Kidhome	int64
Teenhome	int64
Dt_Customer	datetime64[ns]
Recency	int64
MntWines	int64
MntFruits	int64
MntMeatProducts	int64
MntFishProducts	int64
MntSweetProducts	int64
MntGoldProds	int64
NumDealsPurchases	int64
NumWebPurchases	int64
NumCatalogPurchases	int64
NumStorePurchases	int64
NumWebVisitsMonth	int64
AcceptedCmp3	int64
AcceptedCmp4	int64
AcceptedCmp5	int64
AcceptedCmp1	int64
AcceptedCmp2	int64
Complain	int64
Z_CostContact	int64
Z_Revenue	int64
Response	int64
dtype:	object

# Data Cleaning

## Missing Values

Dari 2240 entri data, hanya 24 nilai 'Income' yang hilang, yang berarti hanya sekitar 1% dari data yang memiliki missing values.

**Imputasi dengan Median:** Mengganti 24 nilai yang hilang dengan median agar tidak secara signifikan mengubah distribusi keseluruhan 'Income' dalam dataset. Alasan utamanya adalah karena proporsi missing values yang kecil.

ID	0
Year_Birth	0
Education	0
Marital_Status	0
Income	24
Kidhome	0
Teenhome	0
Dt_Customer	0
Recency	0
MntWines	0
MntFruits	0
MntMeatProducts	0
MntFishProducts	0
MntSweetProducts	0
MntGoldProds	0
NumDealsPurchases	0
NumWebPurchases	0
NumCatalogPurchases	0
NumStorePurchases	0
NumWebVisitsMonth	0
AcceptedCmp3	0
AcceptedCmp4	0
AcceptedCmp5	0
AcceptedCmp1	0
AcceptedCmp2	0
Complain	0
Z_CostContact	0
Z_Revenue	0
Response	0
dtype:	int64



# Data Cleaning

## Outliers

```
{'Year_Birth': 3,  
  'Income': 8,  
  'MntWines': 35,  
  'MntFruits': 227,  
  'MntMeatProducts': 175,  
  'MntFishProducts': 223,  
  'MntSweetProducts': 248,  
  'MntGoldProds': 207}
```

### 1. Demografi Pelanggan:

- `Year_Birth`: Terdapat beberapa outlier dengan jumlah yang relatif kecil, yang mungkin mewakili pelanggan dengan usia yang sangat muda atau tua.
- `Income`: Terdapat outlier dengan jumlah yang juga relatif kecil, yang mungkin menunjukkan adanya pelanggan dengan pendapatan yang sangat tinggi.

### 2. Pengeluaran Produk:

- `MntWines`, `MntMeatProducts`, `MntFishProducts`, `MntSweetProducts`, `MntGoldProds`: Semua menunjukkan adanya outlier, yang mengindikasikan beberapa pelanggan memiliki pengeluaran yang sangat tinggi pada kategori-kategori ini.

### 3. Data Kampanye:

- Untuk kolom `AcceptedCmp1`, `AcceptedCmp2`, `AcceptedCmp3`, `AcceptedCmp4`, `AcceptedCmp5`, dan `Response`, tidak tampak outlier yang signifikan karena data ini bersifat biner (0 atau 1), menggambarkan apakah pelanggan merespon kampanye atau tidak.

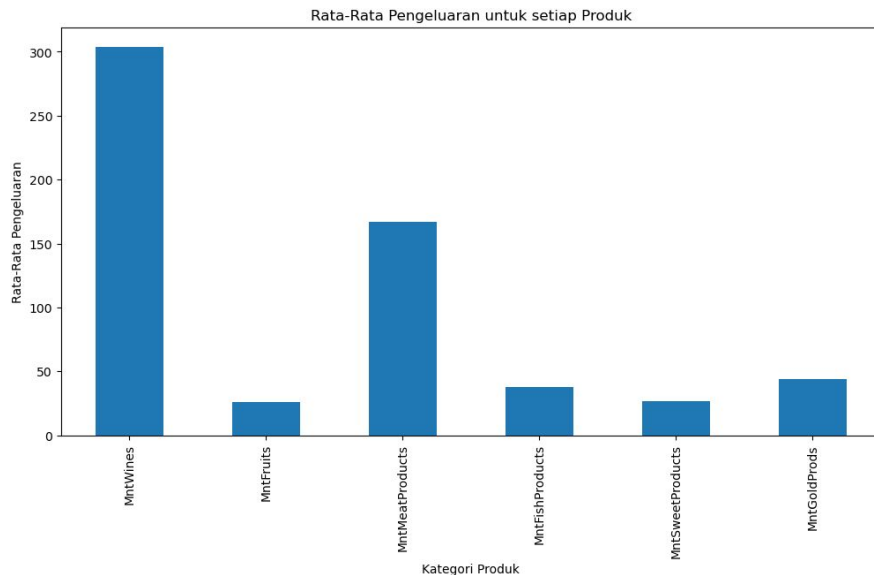
Dalam konteks analisis pola pengeluaran dan efektivitas kampanye, keberadaan outlier ini mungkin menunjukkan adanya segmen pelanggan yang berbeda secara signifikan dalam hal usia, pendapatan, atau pola pengeluaran. Kita mungkin akan bisa memanfaatkan informasi ini jika kita mempertahankan outliers dalam dataset, karena mereka dapat memberikan wawasan penting tentang perilaku pelanggan tertentu.

# Analisa Data

## 02

Visualisasi di samping menunjukkan rata-rata pengeluaran untuk setiap kategori produk. Dari grafik ini, kita dapat melihat bahwa:

- **Anggur (MntWines):** Memiliki pengeluaran rata-rata tertinggi di antara semua kategori, menunjukkan bahwa ini mungkin kategori yang paling populer atau di mana pelanggan cenderung menghabiskan lebih banyak.
- **Produk Daging (MntMeatProducts):** Juga menunjukkan pengeluaran yang signifikan, menunjukkan peminatannya yang tinggi di antara pelanggan.
- **Produk Buah (MntFruits), Ikan (MntFishProducts), Manis (MntSweetProducts), dan Emas (MntGoldProds):** Pengeluaran rata-rata lebih rendah dibandingkan anggur dan daging, yang mungkin menunjukkan bahwa ini adalah kategori tambahan atau sekunder bagi sebagian besar pelanggan.







# Analisis Pengeluaran Produk Berdasarkan Variabel Demografis

Sebelumnya, untuk mempermudah analisis, kita akan mengelompokkan usia dan pendapatan berdasarkan range tertentu. Berikut pengelompokannya:

Usia:

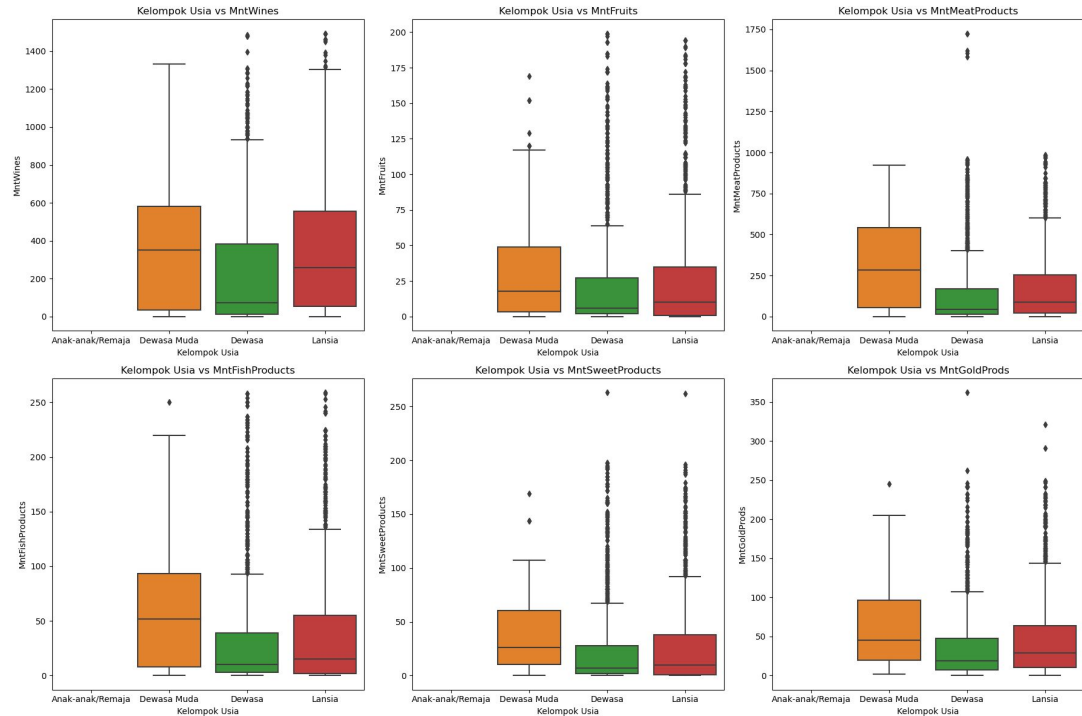
- A. **Anak-anak dan Remaja:** Di bawah 17 tahun.
- B. **Dewasa Muda:** 18-34 tahun.
- C. **Dewasa:** 35-54 tahun.
- D. **Lansia:** 55 tahun ke atas.

Pendapatan:

- **Kuartil Pertama (Q1):** 25% data terendah. Kelompok "Low" mencakup pendapatan dari nilai minimum hingga Q1.
- **Kuartil Kedua (Median, Q2):** Memisahkan data menjadi dua bagian yang sama besar. Kelompok "Medium-Low" mencakup pendapatan dari Q1 hingga Q2.
- **Kuartil Ketiga (Q3):** 75% data terendah. Kelompok "Medium-High" mencakup pendapatan dari Q2 hingga Q3.
- **Kuartil Keempat:** Sisanya hingga nilai maksimum. Kelompok "High" mencakup pendapatan dari Q3 hingga nilai maksimum.

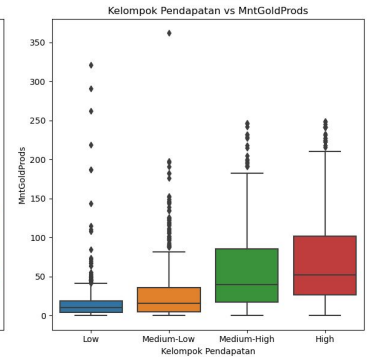
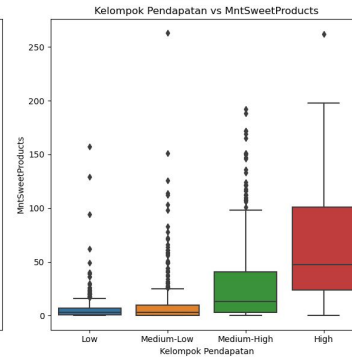
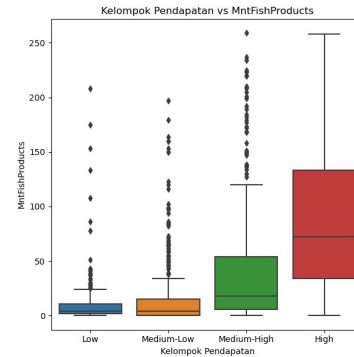
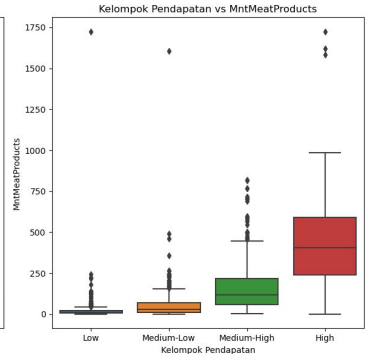
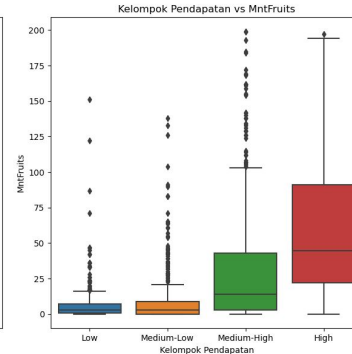
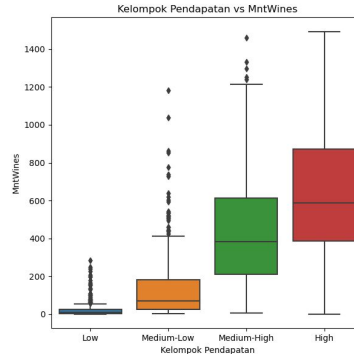
# Analisis Pengeluaran Produk Berdasarkan Variabel Demografis

(Usia vs Pengeluaran  
Produk)



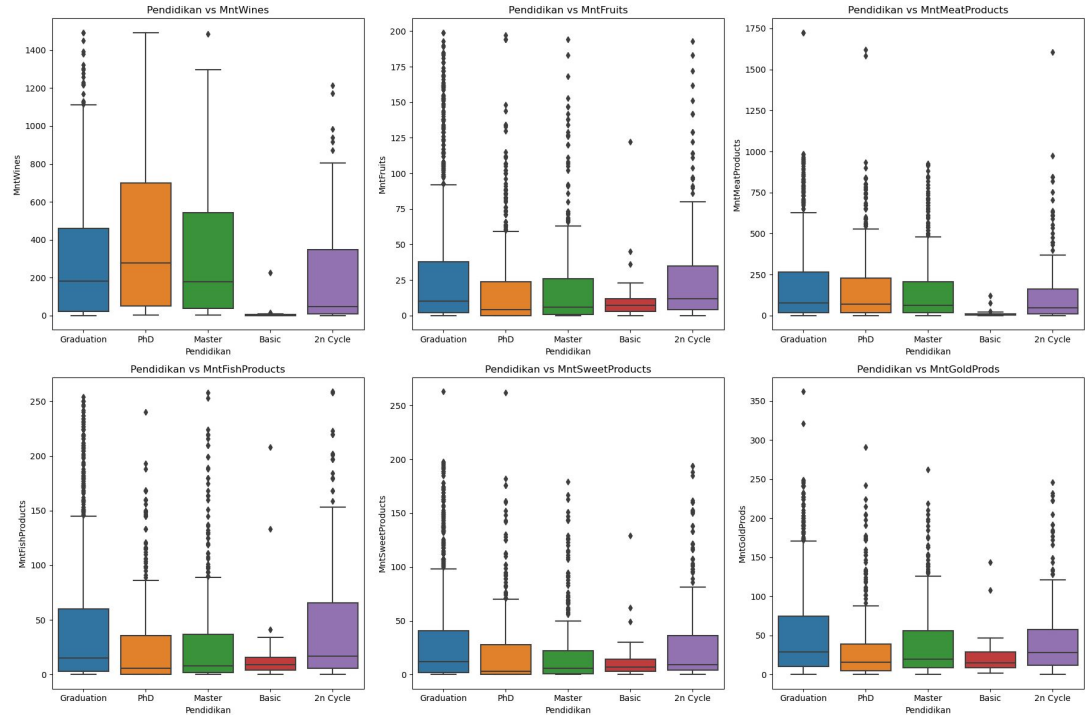
# Analisis Pengeluaran Produk Berdasarkan Variabel Demografis

(Pendapatan vs  
Pengeluaran Produk)



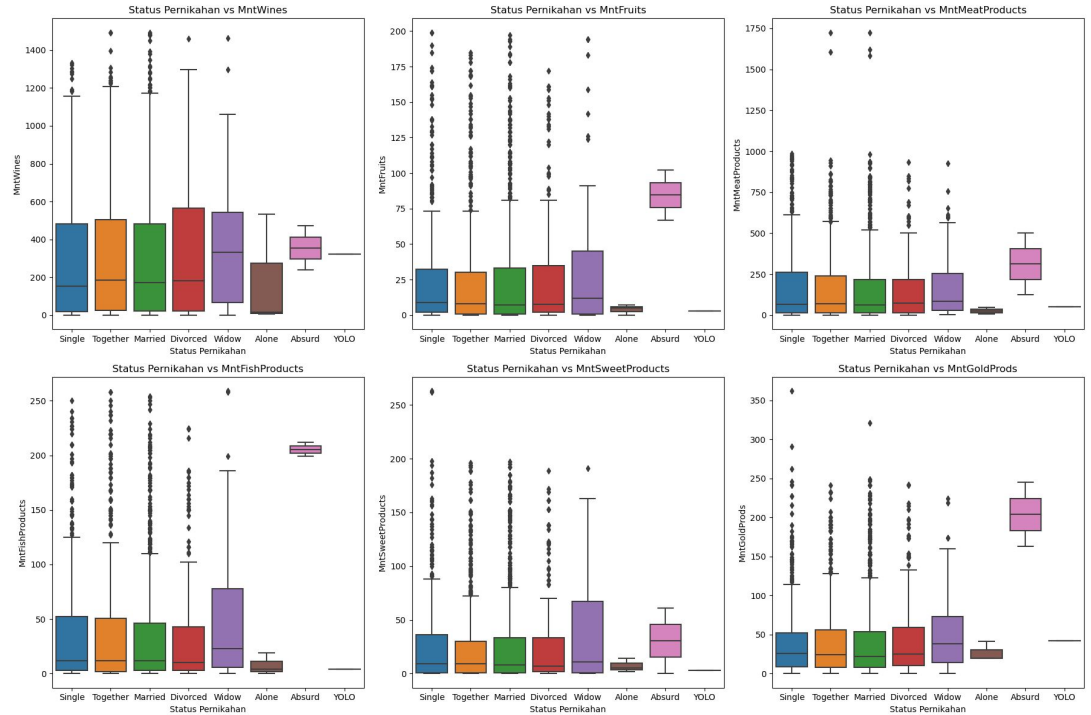
# Analisis Pengeluaran Produk Berdasarkan Variabel Demografis

(Pendidikan vs  
Pengeluaran Produk)



# Analisis Pengeluaran Produk Berdasarkan Variabel Demografis

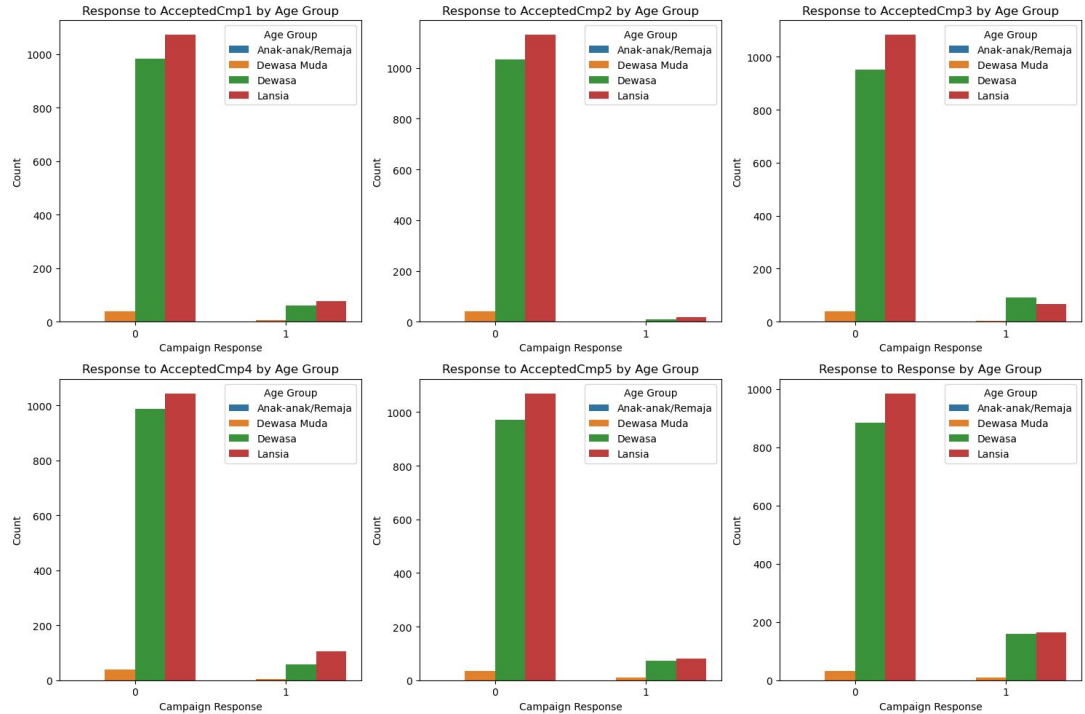
(Status Pernikahan vs  
Pengeluaran Produk)





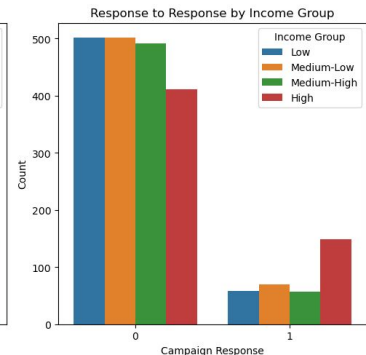
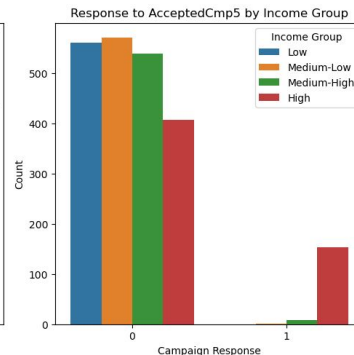
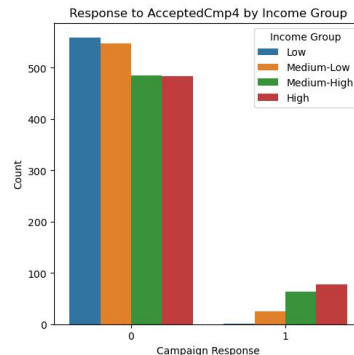
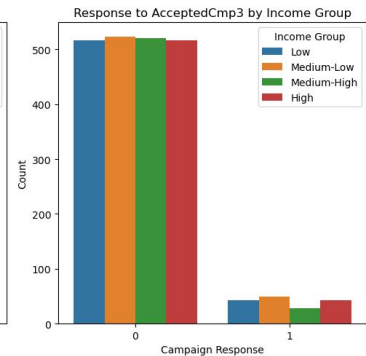
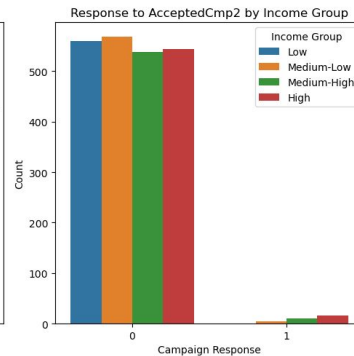
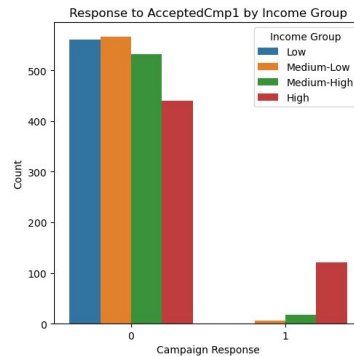
# Analisis Efektivitas Kampanye

(Berdasarkan Usia)



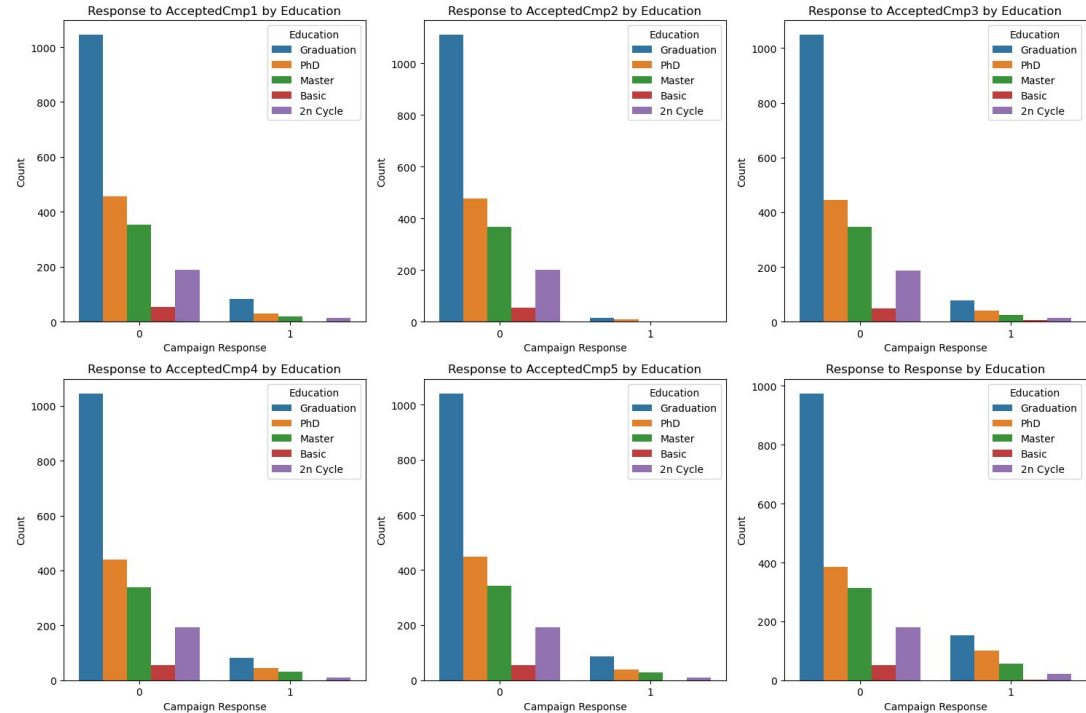
# Analisis Efektivitas Kampanye

(Berdasarkan Pendapatan)



# Analisis Efektivitas Kampanye

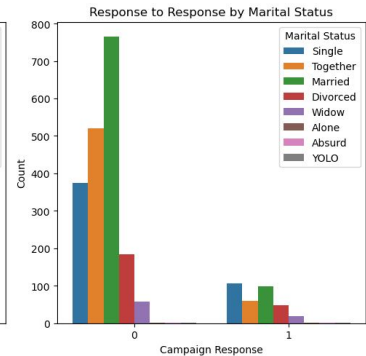
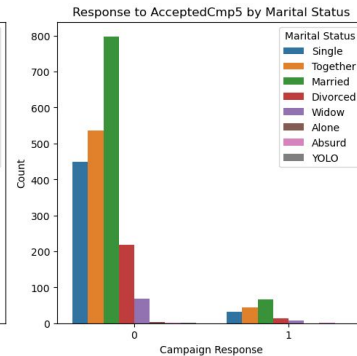
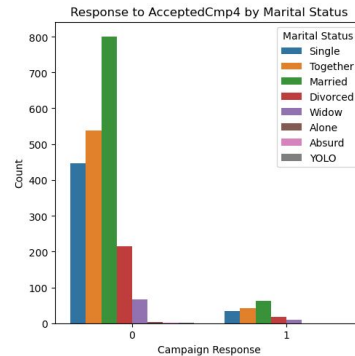
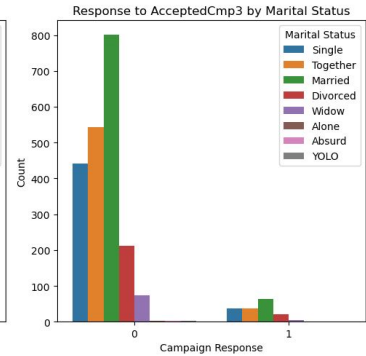
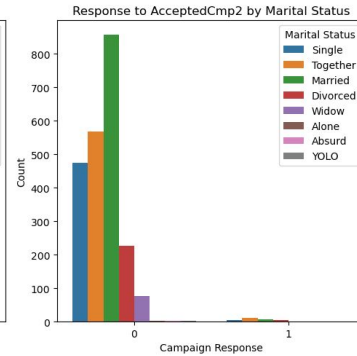
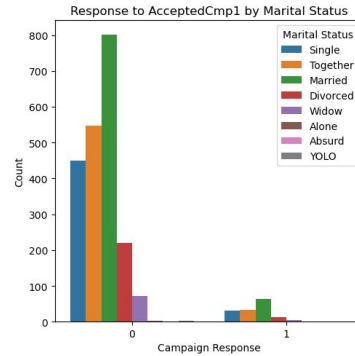
(Berdasarkan Pendidikan)





# Analisis Efektivitas Kampanye

(Berdasarkan Status Pernikahan)





## Kesimpulan

- **Pengeluaran Produk:** Anggur ( `MntWines` ) dan daging ( `MntMeatProducts` ) adalah kategori dengan pengeluaran tertinggi. Ini menunjukkan preferensi kuat pelanggan terhadap produk ini.
- **Demografi:** Ada perbedaan signifikan dalam pola pengeluaran berdasarkan demografi. Pendapatan, tingkat pendidikan, dan status pernikahan mempengaruhi pengeluaran pada kategori produk tertentu.
- **Efektivitas Kampanye:** Tingkat respons terhadap kampanye pemasaran cenderung rendah, dengan kampanye terakhir ( `Response` ) menunjukkan efektivitas yang lebih tinggi. Pendapatan dan pendidikan berperan dalam menentukan responsivitas terhadap kampanye.

### 1. Analisis Pola Pengeluaran:

- Ada perbedaan signifikan dalam rata-rata pengeluaran pada produk tertentu antara berbagai kelompok demografi. Kelompok dengan pendapatan lebih tinggi, pendidikan tinggi (PhD), dan yang menikah cenderung menghabiskan lebih pada kategori tertentu seperti anggur dan daging.
- Distribusi pengeluaran pada produk tertentu berbeda antara kelompok demografi. Anggur dan daging menonjol sebagai kategori dengan pengeluaran tertinggi, terutama di antara kelompok pendapatan dan pendidikan yang lebih tinggi.

### 2. Efektivitas Kampanye:

- Respons terhadap kampanye pemasaran bervariasi berdasarkan demografi. Kelompok pendapatan tinggi dan pelanggan dengan pendidikan PhD lebih responsif terhadap kampanye.
- Ada korelasi antara tingkat keberhasilan kampanye dan karakteristik demografi seperti pendapatan dan pendidikan. Pelanggan dengan pendapatan dan pendidikan lebih tinggi cenderung lebih responsif.





## Rekomendasi

### 1. Fokus pada Segmen Berpendapatan Tinggi dan Berpendidikan Tinggi:

- Tingkatkan penargetan pemasaran pada pelanggan dengan pendapatan tinggi dan pendidikan PhD, terutama untuk produk premium seperti anggur dan daging.

### 2. Kampanye Pemasaran yang Lebih Personal dan Segmentasi Pasar:

- Kembangkan kampanye yang lebih disesuaikan berdasarkan demografi pelanggan. Misalnya, penawaran khusus untuk pelanggan yang menikah atau berpendidikan tinggi.

### 3. Analisis dan Optimasi Kampanye Berkelanjutan:

- Evaluasi dan sesuaikan kampanye berdasarkan data respons pelanggan. Manfaatkan kampanye yang berhasil di masa lalu sebagai referensi untuk kampanye mendatang.

### 4. Penyesuaian Produk dan Penawaran:

- Pertimbangkan penyesuaian stok dan penawaran produk berdasarkan preferensi kelompok pendapatan dan pendidikan. Misalnya, meningkatkan stok anggur dan daging berkualitas tinggi di lokasi dengan pelanggan berpendapatan tinggi.

### 5. Strategi Promosi Berdasarkan Usia dan Status Pernikahan:

- Kembangkan promosi yang menargetkan kelompok usia dewasa dan lansia, serta pelanggan yang menikah, karena mereka cenderung menghabiskan lebih pada kategori tertentu seperti anggur dan daging.



# Terima kasih.

