



*Politeknik Elektronika Negeri
Surabaya*

Project Mid-Semester

Ekonometrika Terapan

Diampu Oleh: Ronny Susetyoko S.Si., M.Si

Elza Hiya Nadhifa 3323600040



*Politeknik Elektronika Negeri
Surabaya*

Title Project

Analisis Faktor-faktor yang Memengaruhi Penggunaan Diskon pada
Platform E-Commerce berdasarkan Customer Behavior
Menggunakan Model Regresi Probit



Background Study Case

cs on the value of e-commerce sales by businesses. It benefits from release, by Eurostat, of figures for many EU and partner countries. merce sales were generated by businesses across 43 developed ers of worldwide GDP. This represents a 15 per cent increase over risen a further 10 per cent - to almost \$27 trillion - in 2022.

ough e-com
all cases, th
e of busines

67 Coupon Statistics 2025 — Redemption Rates & Usage

By Shubham Singh / August 20, 2025

In 2025, the global digital coupon market is valued at \$10.6 billion, up from \$8.96 billion in 2024 and projected to exceed \$34.43 billion by 2032. Nearly 90% of U.S. consumers use coupons, with 62% actively searching for promo codes online.

BPS Sebut Konsumsi Masyarakat Beralih ke Online, Daya Beli Pulih?

BPS melaporkan konsumsi rumah tangga tumbuh 5,12% di Q2 2025, didorong oleh peralihan belanja ke online. PDB Indonesia tumbuh 5,12% YoY, melebihi proyeksi ekonom.

Dalam beberapa tahun terakhir, dunia mengalami perubahan besar dalam pola konsumsi masyarakat. Aktivitas belanja yang sebelumnya banyak dilakukan secara langsung di toko kini bergeser ke platform digital seiring pesatnya pertumbuhan e-commerce global. Berdasarkan laporan United Nations Conference on Trade and Development (UNCTAD, 2023), nilai transaksi e-commerce dunia mencapai sekitar US\$27 triliun pada tahun 2022, dengan sebagian besar transaksi terjadi di pasar domestik. Pertumbuhan ini dipicu oleh kemajuan teknologi digital, peningkatan konektivitas internet, serta perubahan gaya hidup masyarakat yang semakin mengutamakan efisiensi dan kenyamanan dalam berbelanja.

Perubahan perilaku belanja tersebut turut mengubah strategi perusahaan dalam menarik minat konsumen. Dalam persaingan pasar digital yang semakin ketat, pemberian diskon dan promo harga menjadi salah satu strategi pemasaran yang paling sering digunakan untuk mendorong keputusan pembelian. Berdasarkan laporan DemandSage (2025), sekitar 62% konsumen secara aktif mencari kode promo atau kupon saat berbelanja online, dan tren ini terus meningkat dari tahun ke tahun. Hal ini menunjukkan bahwa penggunaan diskon telah menjadi bagian penting dari perilaku konsumen digital modern, di mana konsumen semakin rasional dan sensitif terhadap harga.

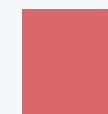
The Problem is?



Meskipun banyak customer cenderung mencari potongan harga, tidak semua customer merespons diskon dengan cara yang sama. Ada customer yang sangat sensitif terhadap harga dan akan berbelanja hanya saat ada promo, sementara sebagian lainnya tetap berbelanja meskipun tanpa potongan harga. Perbedaan perilaku ini penting untuk dipahami karena strategi diskon tidak selalu efektif bagi semua segmen konsumen. Jika perusahaan memberikan diskon tanpa memahami perilaku customer, hal itu bisa menurunkan margin keuntungan tanpa memberikan dampak terhadap penjualan atau loyalitas pelanggan.



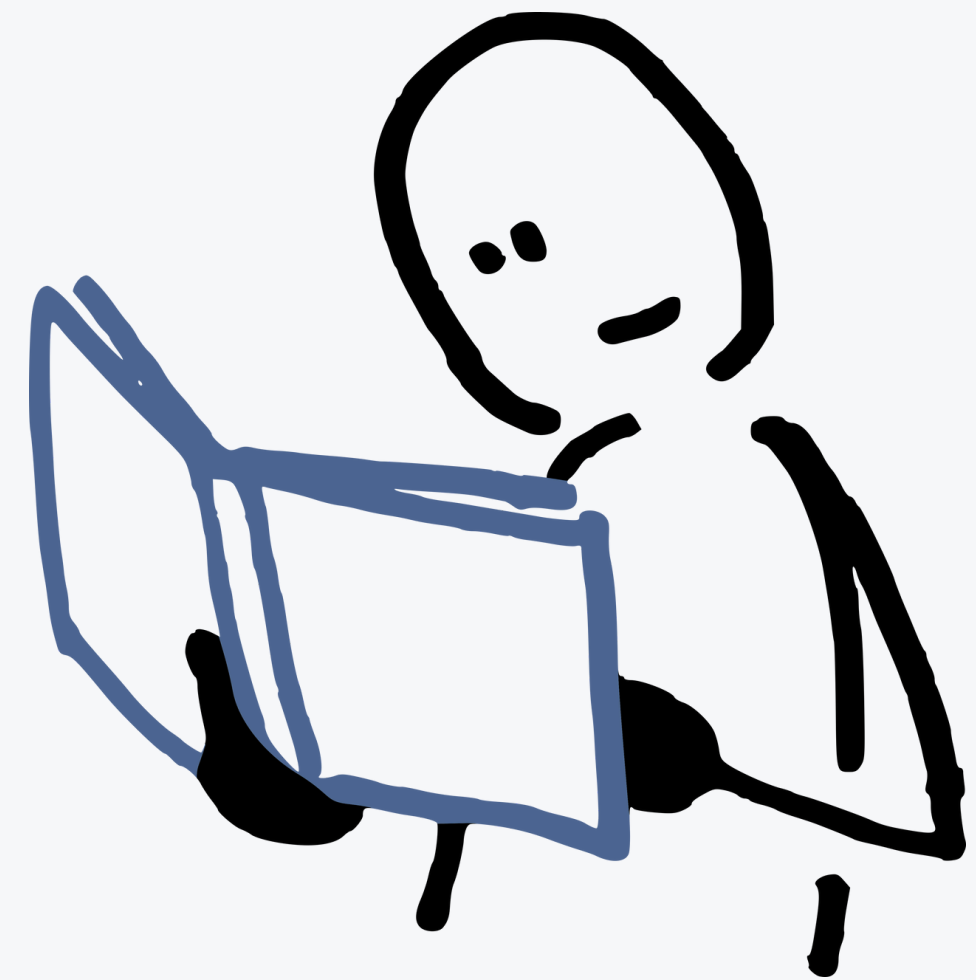
Masalah yang terjadi



Akibat yang akan terjadi

The Solution?

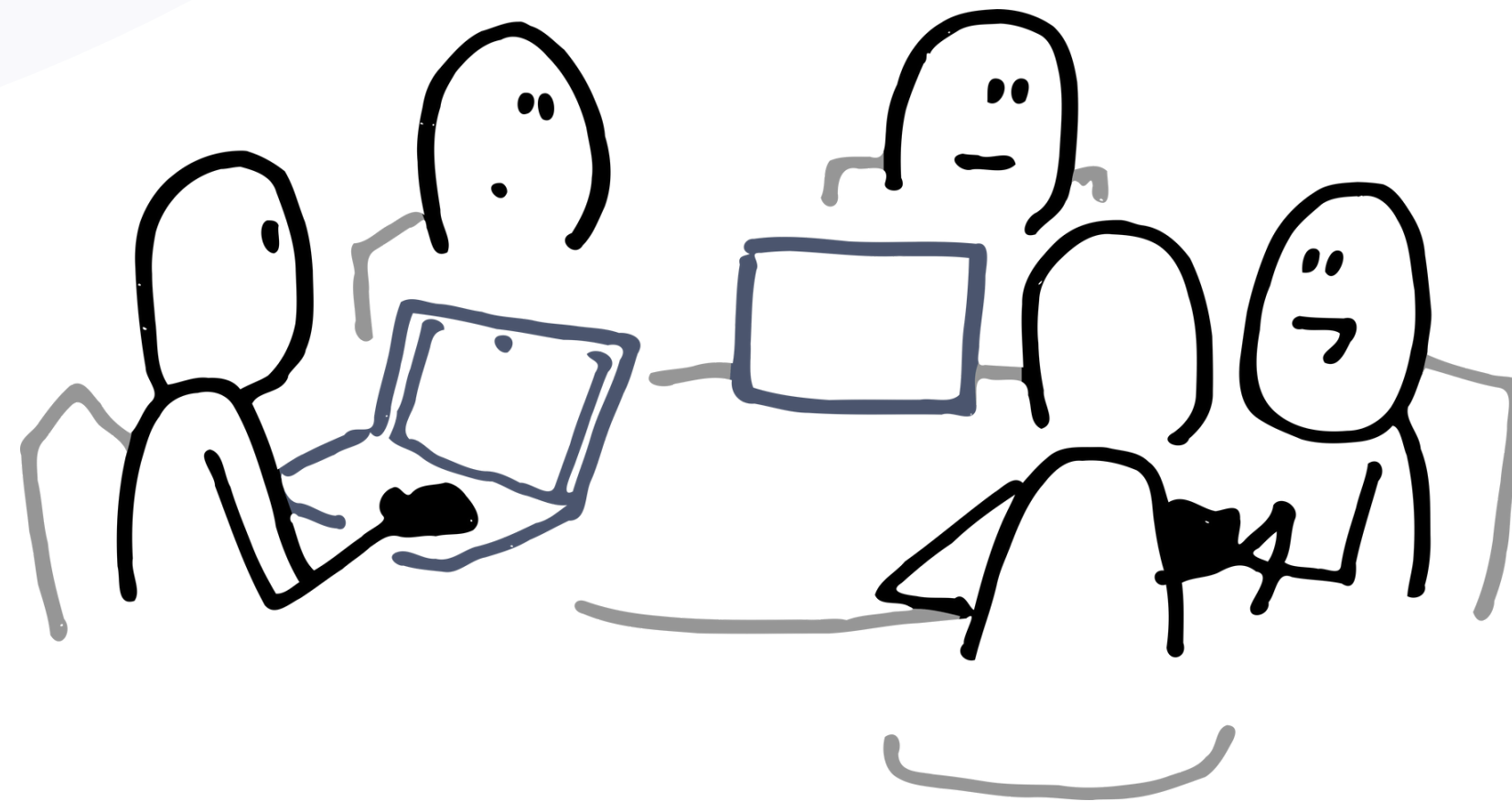
Untuk memahami perbedaan perilaku customer dalam merespons diskon, penelitian ini akan melakukan analisis faktor-faktor yang memengaruhi penggunaan diskon pada platform e-commerce menggunakan model regresi Probit. Penelitian akan menganalisis variabel-variabel customer behavior untuk melihat mana yang paling berpengaruh terhadap keputusan menggunakan diskon. Hasilnya akan membantu perusahaan memahami segmen konsumen yang lebih responsif terhadap promo, sehingga strategi diskon dapat dibuat lebih tepat sasaran dan efisien.





*Politeknik Elektronika Negeri
Surabaya*

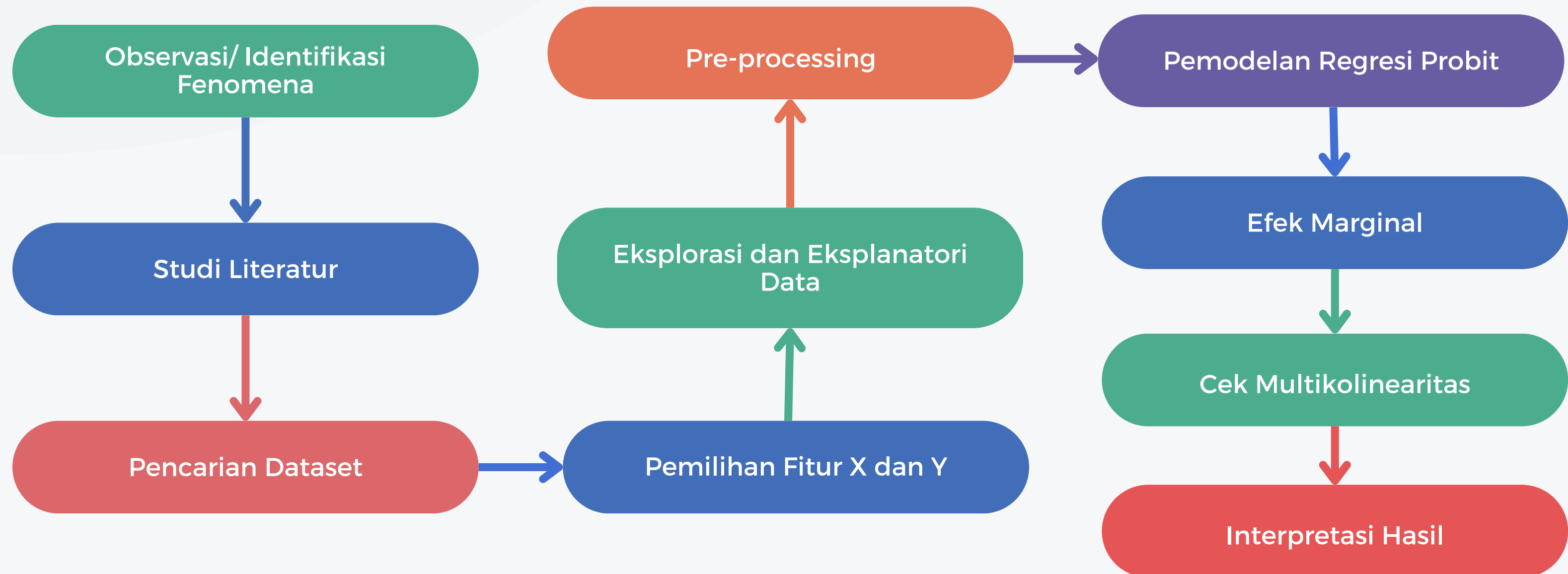
Discussion



Next Slide -->



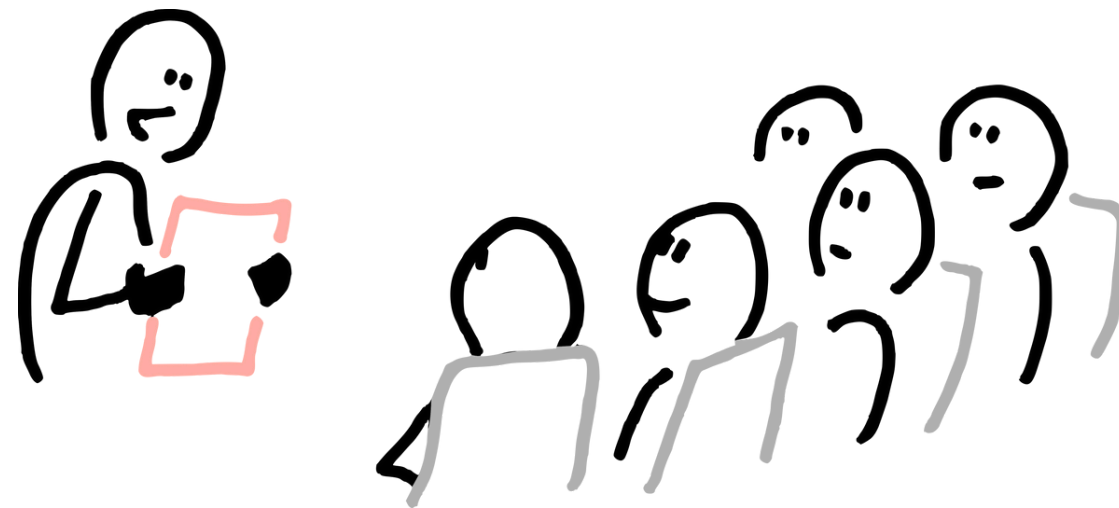
Flowchart of the research process





*Politeknik Elektronika Negeri
Surabaya*

Link Source Code



Klik: [Source Code](#)

Next Slide -->



Dataset

Nama Variabel	Deskripsi
Customer_ID	Unique identifier for each customer.
Age	Customer's age (integer).
Gender	Customer's gender (categorical: Male, Female, Non-binary, Other).
Income_Level	Customer's income level (categorical: Low, Middle, High).
Marital_Status	Customer's marital status (categorical: Single, Married, Divorced, Widowed).
Education_Level	Highest level of education completed (categorical: High School, Bachelor's, Master's, Doctorate).
Occupation	Customer's occupation (categorical: Various job titles).
Location	Customer's location (city, region, or country).
Purchase_Category	Category of purchased products (e.g., Electronics, Clothing, Groceries).
Purchase_Amount	Amount spent during the purchase (decimal).
Frequency_of_Purchase	Number of purchases made per month (integer).
Purchase_Channel	The purchase method (categorical: Online, In-Store, Mixed).
Brand_Loyalty	Loyalty to brands (1-5 scale).
Product_Rating	Rating given by the customer to a purchased product (1-5 scale).
Time_Spent_on_Product_Research	Time spent researching a product (integer, hours or minutes).

Social_Media_Influence	Influence of social media on purchasing decision (categorical: High, Medium, Low, None).
Discount_Sensitivity	Sensitivity to discounts (categorical: Very Sensitive, Somewhat Sensitive, Not Sensitive).
Return_Rate	Percentage of products returned (decimal).
Customer_Satisfaction	Overall satisfaction with the purchase (1-10 scale).
Engagement_with_Ads	Engagement level with advertisements (categorical: High, Medium, Low, None).
Device_Used_for_Shopping	Device used for shopping (categorical: Smartphone, Desktop, Tablet).
Payment_Method	Method of payment used for the purchase (categorical: Credit Card, Debit Card, PayPal, Cash, Other).
Time_of_Purchase	Timestamp of when the purchase was made (date/time).
Discount_Used	Whether the customer used a discount (Boolean: True/False).
Customer_Loyalty_Program_Member	Whether the customer is part of a loyalty program (Boolean: True/False).
Purchase_Intent	The intent behind the purchase (categorical: Impulsive, Planned, Need-based, Wants-based).
Shipping_Preference	Shipping preference (categorical: Standard, Express, No Preference).
Payment_Frequency	Frequency of payment (categorical: One-time, Subscription, Installments).

Dataset yang digunakan merupakan data ecommerce customer behavior yang diperoleh secara daring melalui website Kaggle



Dependen and Independen Features

Fitur Dependen

'Discount_Used'

'Age', 'Gender', 'Income_Level', 'Marital_Status', 'Education_Level',
'Occupation', 'Purchase_Category', 'Purchase_Amount', 'Frequency_of_Purchase',
'Purchase_Channel', 'Brand_Loyalty', 'Product_Rating',
'Time_Spent_on_Product_Research(hours)', 'Social_Media_Influence',
'Discount_Sensitivity', 'Return_Rate', 'Customer_Satisfaction',
'Engagement_with_Ads', 'Device_Used_for_Shopping', 'Payment_Method',
'Customer_Loyalty_Program_Member', 'Purchase_Intent',
'Shipping_Preference', 'Time_to_Decision'

Fitur Independen



EDA

Informasi Dataset

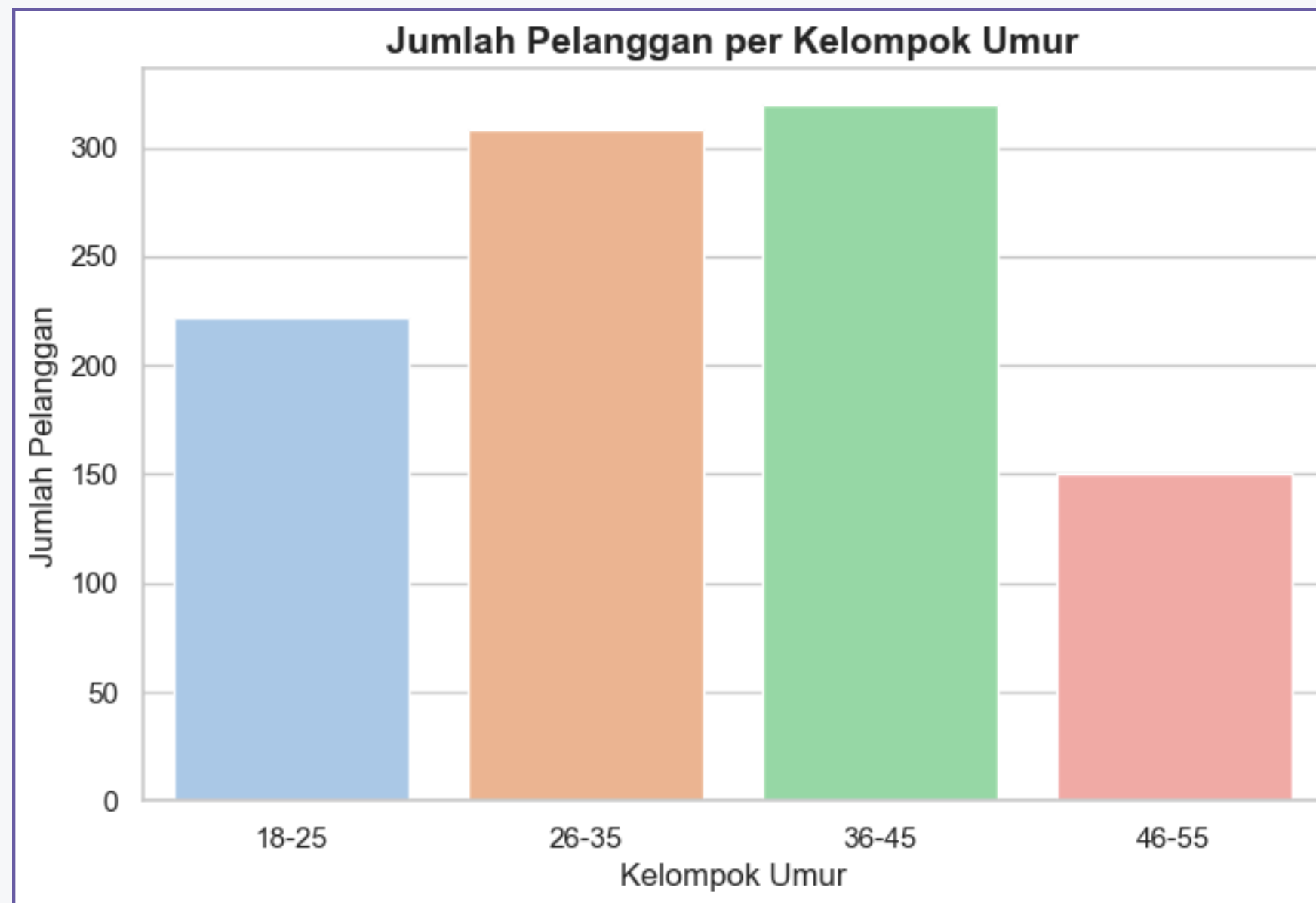
Terlihat bahwa pada dataset terdapat data NaN atau missing value yang perlu untuk dilakukan handling serta mayoritas tipe data variabel masih berbentuk object sehingga perlu untuk dilakukan handling ubah tipe data atau diencoding

0	Customer_ID	1000	non-null	object
1	Age	1000	non-null	int64
2	Gender	1000	non-null	object
3	Income_Level	1000	non-null	object
4	Marital_Status	1000	non-null	object
5	Education_Level	1000	non-null	object
6	Occupation	1000	non-null	object
7	Location	1000	non-null	object
8	Purchase_Category	1000	non-null	object
9	Purchase_Amount	1000	non-null	object
10	Frequency_of_Purchase	1000	non-null	int64
11	Purchase_Channel	1000	non-null	object
12	Brand_Loyalty	1000	non-null	int64
13	Product_Rating	1000	non-null	int64
14	Time_Spent_on_Product_Research(hours)	1000	non-null	float64
15	Social_Media_Influence	753	non-null	object
16	Discount_Sensitivity	1000	non-null	object
17	Return_Rate	1000	non-null	int64
18	Customer_Satisfaction	1000	non-null	int64
19	Engagement_with_Ads	744	non-null	object
20	Device_Used_for_Shopping	1000	non-null	object
21	Payment_Method	1000	non-null	object
22	Time_of_Purchase	1000	non-null	object
23	Discount_Used	1000	non-null	bool
24	Customer_Loyalty_Program_Member	1000	non-null	bool
25	Purchase_Intent	1000	non-null	object
26	Shipping_Preference	1000	non-null	object
27	Time to Decision	1000	non-null	int64



Dataset Explanatory

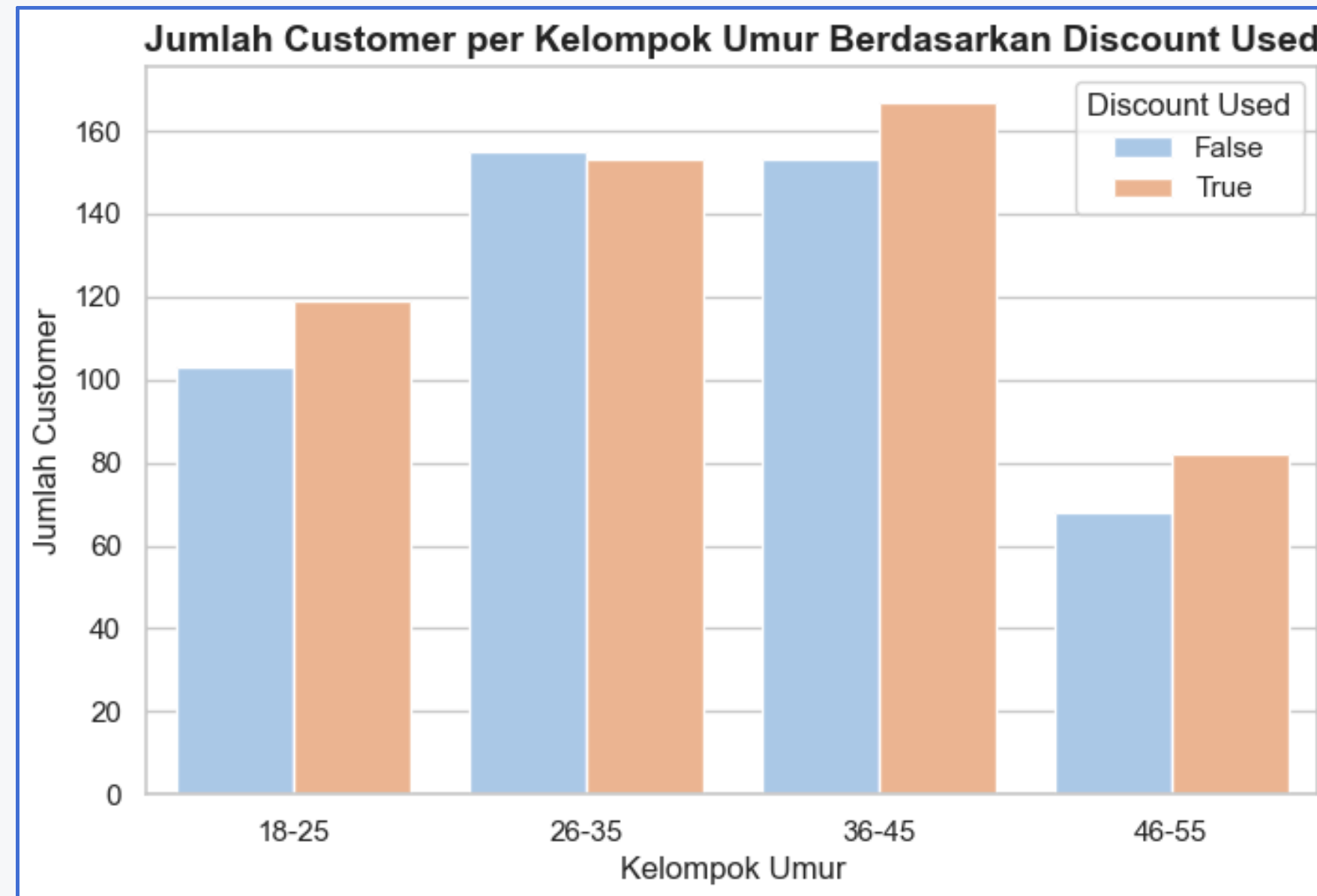
Distribusi Umur



Berdasarkan analisis distribusi umur pada dataset, sebagian besar customer berada dalam rentang usia 26–45 tahun, dengan kelompok 35–45 tahun menjadi yang terbesar, diikuti oleh kelompok 26–35 tahun. Kelompok usia muda (18–25 tahun) dan kelompok usia lebih tua (46–55 tahun) relatif lebih sedikit. Hal ini menunjukkan bahwa segmen pelanggan utama yang aktif bertransaksi di platform e-commerce berada pada usia dewasa muda hingga paruh baya.

Dataset Explanatory

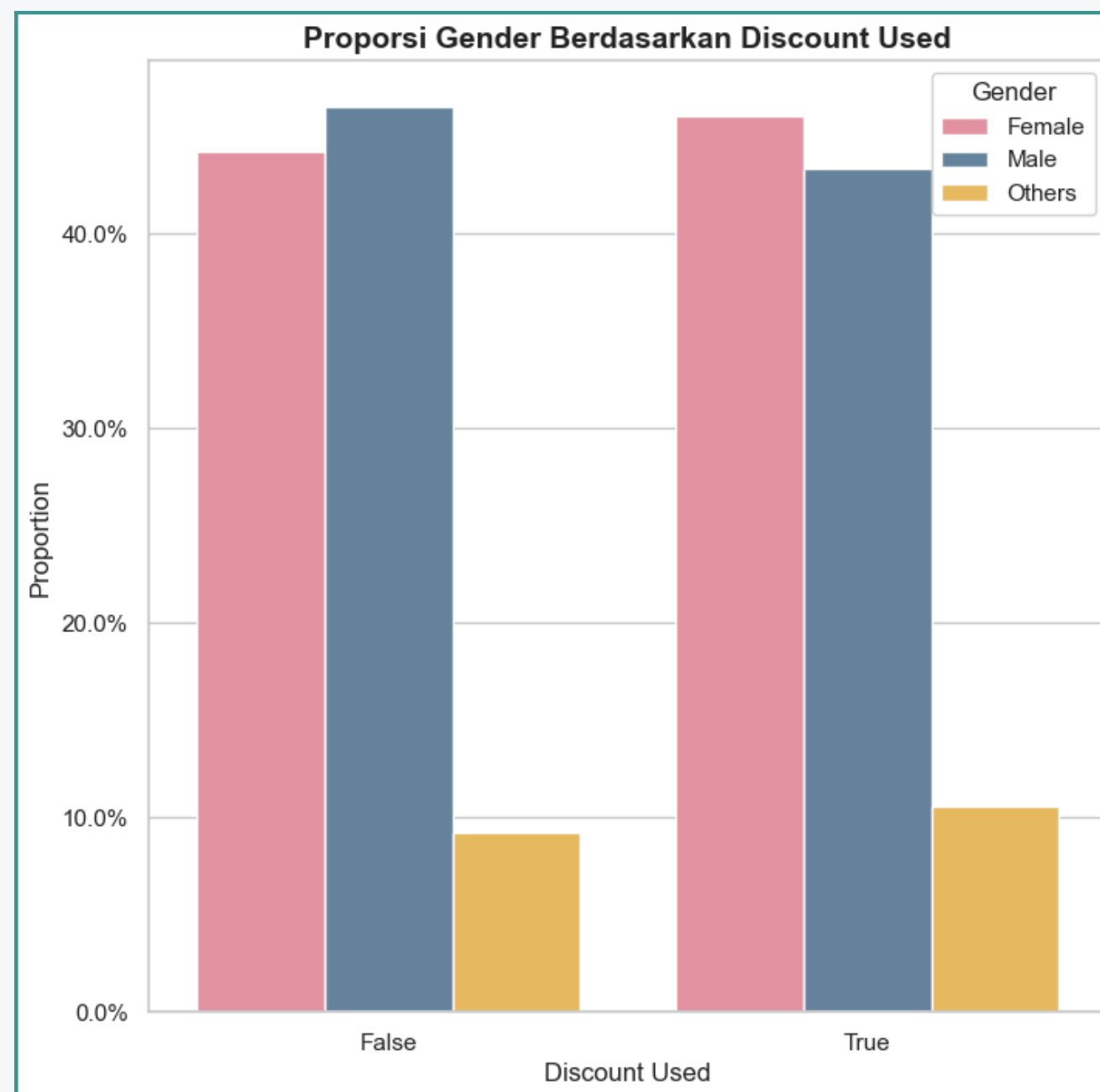
Jumlah Customer per Kelompok Umur berdasarkan Discount Used



Pada kelompok usia 26–35 tahun, jumlah customer yang tidak menggunakan diskon (False) sedikit lebih banyak dibandingkan yang menggunakan diskon, menunjukkan bahwa segmen usia ini cenderung bertransaksi tanpa memanfaatkan promo. Sebaliknya, pada kelompok usia 36–45 tahun, 18–25 tahun, dan 46–55 tahun, pengguna diskon lebih dominan, menandakan bahwa pelanggan pada usia paruh baya maupun kelompok usia muda dan tua lebih memilih memanfaatkan penggunaan discount

Dataset Explanatory

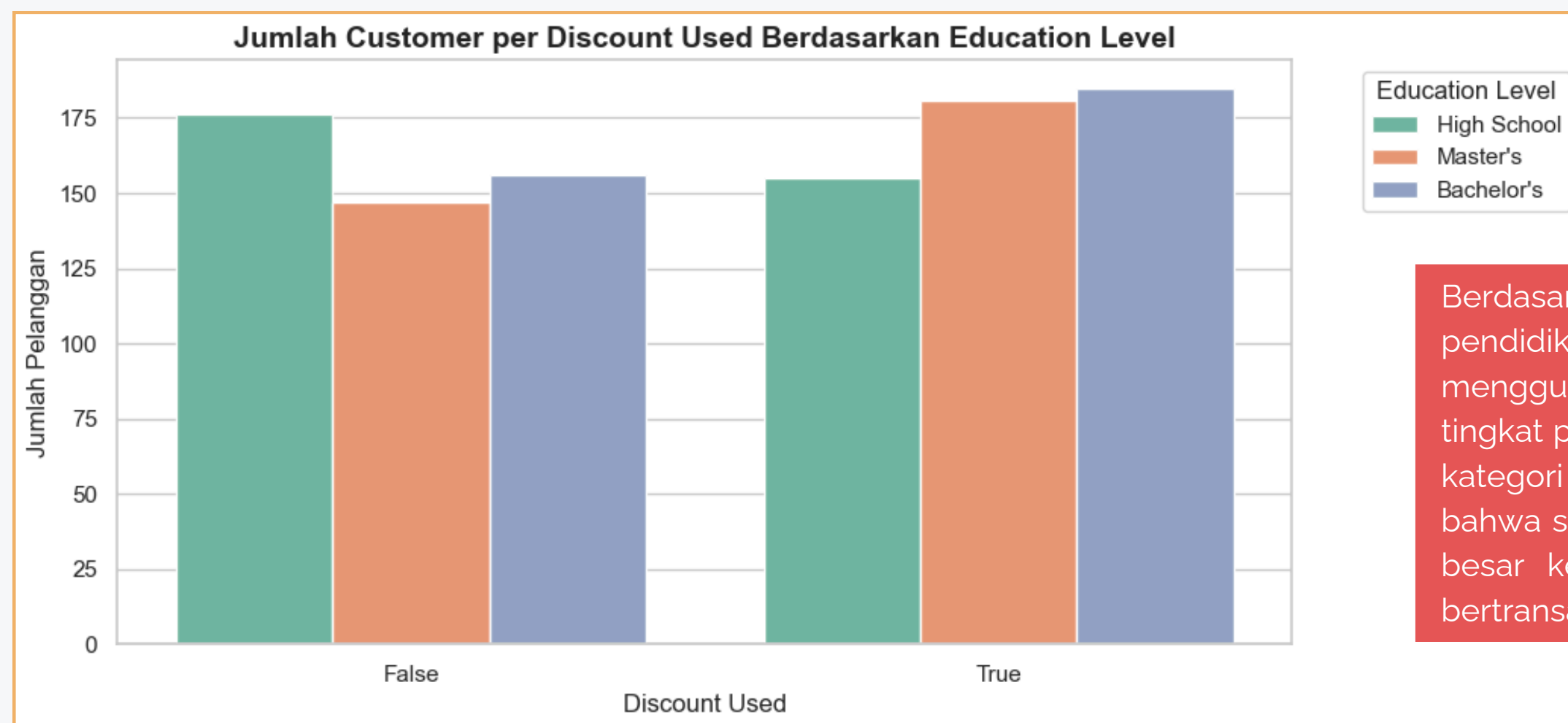
Proporsi Gender Berdasarkan Discount Used



Berdasarkan grafik, terlihat bahwa gender Male mendominasi pada kategori tidak menggunakan diskon (False), sedangkan gender Female lebih banyak pada kategori menggunakan diskon (True). Hal ini menunjukkan bahwa laki-laki cenderung bertransaksi tanpa memanfaatkan promo, sementara perempuan lebih responsif terhadap penawaran diskon.

Dataset Explanatory

Jumlah Customer per Discount Used Berdasarkan Education Level

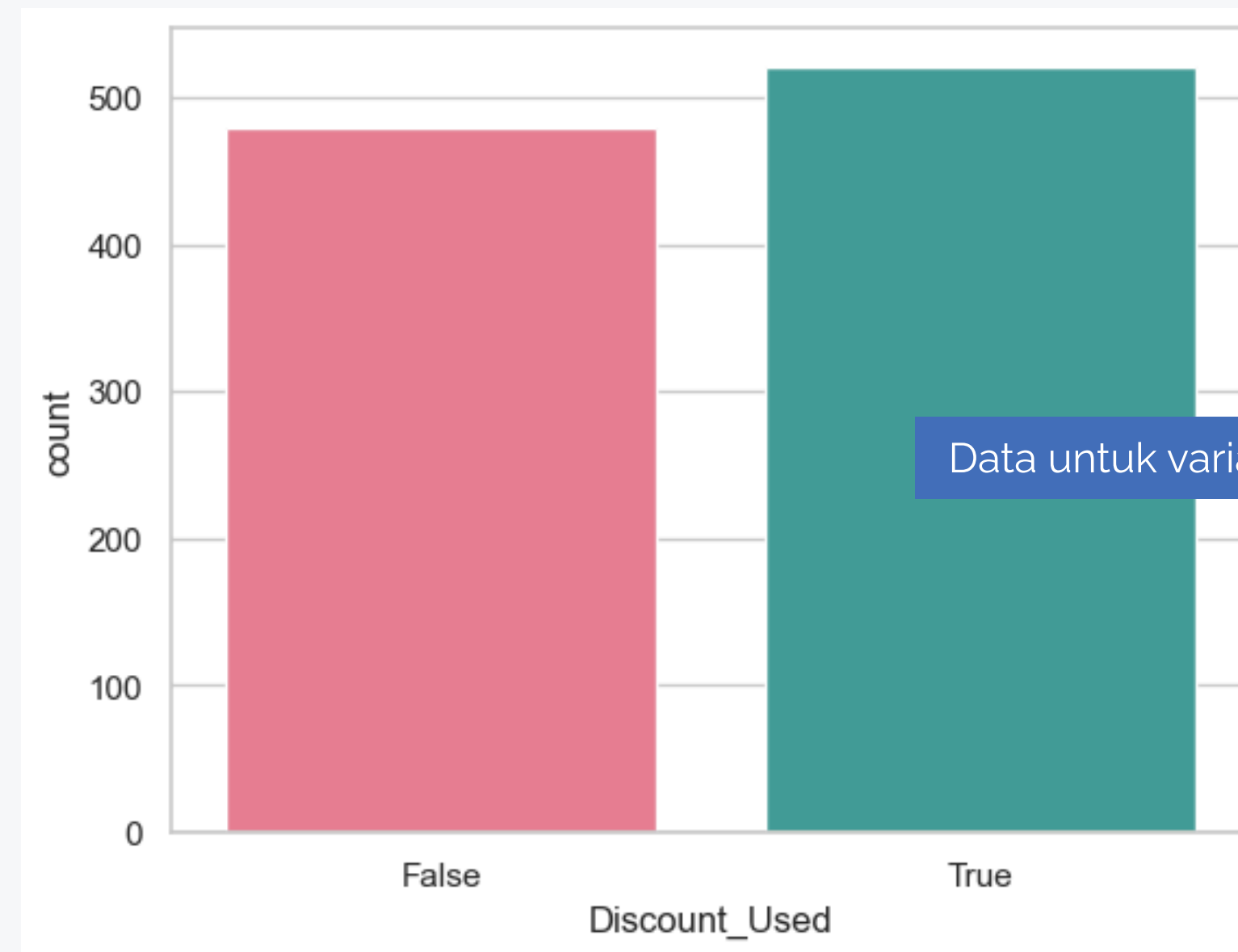


Berdasarkan grafik, terlihat bahwa pelanggan dengan tingkat pendidikan High School mendominasi pada kategori tidak menggunakan diskon (False), sedangkan pelanggan dengan tingkat pendidikan Bachelor's dan Master's lebih banyak pada kategori menggunakan diskon (True). Hal ini menunjukkan bahwa semakin tinggi tingkat pendidikan seseorang, semakin besar kemungkinan mereka memanfaatkan diskon dalam bertransaksi.



Check Imbalanced Dataset

cek apakah data sudah balance atau belum pada variabel dependen (y)



Data untuk variabel dependen sudah balance

Pre-processing

Drop Kolom

Beberapa variabel seperti Customer_ID, Location, Age_Group, dan Time_of_Purchase dihapus dari dataset karena dianggap tidak relevan secara langsung terhadap analisis yang dilakukan

Handling Missing Value

Melakukan proses imputasi untuk mengisi nilai NaN pada kolom Social_Media_Influence dan Engagement_with_Ads. Pengisian dilakukan berdasarkan kelompok Device_Used_for_Shopping, di mana setiap nilai kosong diisi dengan kategori yang paling sering muncul (modus) pada Device_Used_for_Shopping

Ubah Tipe Dataset

Melakukan pembersihan data pada kolom Purchase_Amount dengan menghilangkan tanda \$ dan karakter koma menggunakan fungsi replace, kemudian mengubah tipe datanya menjadi float agar dapat digunakan dalam analisis numerik.

Encoding

Pertama, kolom Discount_Used dan Customer_Loyalty_Program_Member dikonversi menjadi tipe integer (0 dan 1) karena bersifat biner. Selanjutnya, dilakukan ordinal encoding untuk beberapa variabel yang memiliki urutan logis, seperti Social_Media_Influence, Engagement_with_Ads, Income_Level, Occupation, Discount_Sensitivity, dan Education_Level, dengan memberikan nilai numerik sesuai tingkatannya (next slide)

Pre-processing

Encoding (2)

Setelah itu, variabel kategorikal lainnya seperti Gender, Marital_Status, Purchase_Category, Purchase_Channel, Device_Used_for_Shopping, Payment_Method, Purchase_Intent, dan Shipping_Preference diubah menggunakan Label Encoding dengan bantuan LabelEncoder()

Normalisasi

Melakukan proses normalisasi data menggunakan metode Min-Max Scaler pada variabel X (Independen) agar seluruh nilai fitur berada pada rentang 0 hingga 1.



Modeling

Regresi Probit

Optimization terminated successfully.
Current function value: 0.674814
Iterations 4

Probit Regression Results

=====

Dep. Variable:	Discount_Used	No. Observations:	1000
Model:	Probit	Df Residuals:	976
Method:	MLE	Df Model:	23
Date:	Tue, 07 Oct 2025	Pseudo R-squ.:	0.02521
Time:	13:40:55	Log-Likelihood:	-674.81
converged:	True	LL-Null:	-692.26
Covariance Type:	nonrobust	LLR p-value:	0.05320

=====

	coef	std err	z	P> z	[0.025	0.975]
Age	0.0254	0.133	0.191	0.848	-0.235	0.286
Gender	-0.0253	0.123	-0.206	0.837	-0.265	0.215
Income_Level	-0.1474	0.080	-1.848	0.065	-0.304	0.009
Marital_Status	0.0384	0.105	0.367	0.714	-0.167	0.244
Education_Level	0.2039	0.098	2.091	0.037	0.013	0.395
Occupation	-0.0640	0.081	-0.794	0.427	-0.222	0.094
Purchase_Category	-0.0340	0.132	-0.257	0.797	-0.293	0.225
Purchase_Amount	-0.0621	0.135	-0.460	0.645	-0.327	0.202
Frequency_of_Purchase	0.1534	0.124	1.236	0.216	-0.090	0.397
Purchase_Channel	0.0005	0.098	0.005	0.996	-0.192	0.193
Brand_Loyalty	-0.2404	0.111	-2.165	0.030	-0.458	-0.023
Product_Rating	-0.2759	0.110	-2.515	0.012	-0.491	-0.061
Time_Spent_on_Product_Research(hours)	-0.0044	0.100	-0.044	0.965	-0.200	0.192
Social_Media_Influence	-0.0436	0.096	-0.453	0.651	-0.232	0.145
Discount_Sensitivity	0.0521	0.097	0.540	0.589	-0.137	0.242

=====

Hasil modeling menggunakan regresi probit

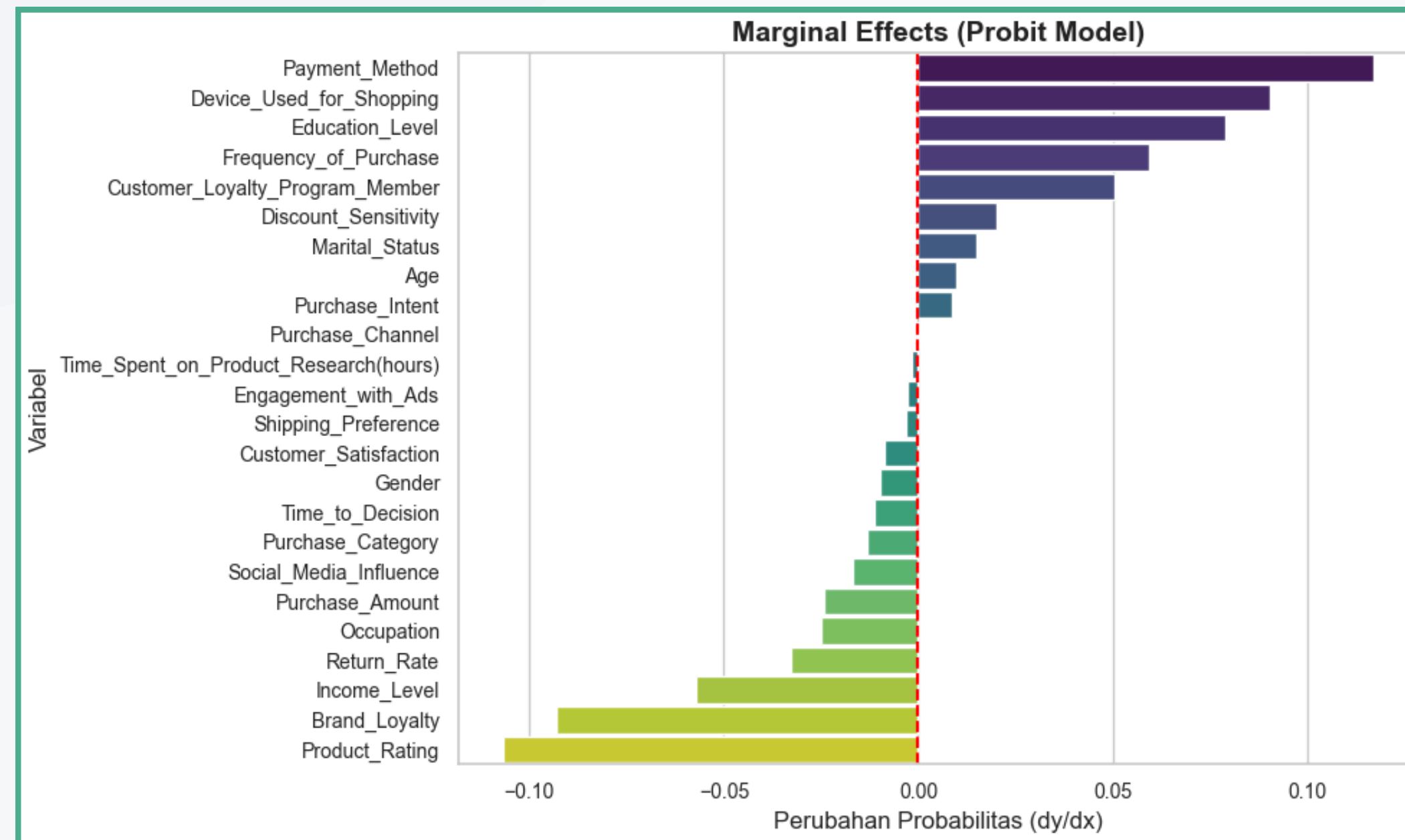


Modeling

Filtering variabel p-value < 0.05

	Variable	Coef	Std.Err	z	P> z	[0.025	0.975]
4	Education_Level	0.203945	0.097543	2.090819	0.036544	0.012764	0.395125
10	Brand_Loyalty	-0.240375	0.111048	-2.164610	0.030418	-0.458024	-0.022725
11	Product_Rating	-0.275928	0.109712	-2.515011	0.011903	-0.490960	-0.060895
18	Device_Used_for_Shopping	0.233239	0.105412	2.212647	0.026922	0.026636	0.439843
19	Payment_Method	0.302177	0.112473	2.686668	0.007217	0.081734	0.522619

Marginal Effects





Multikolinearitas Check

Checking multikolinearitas with VIF

	Variable	VIF
0	const	89.685366
1	Age	1.016096
2	Gender	1.012936
3	Income_Level	1.022512
4	Marital_Status	1.013201
5	Education_Level	1.017998
6	Occupation	1.035802
7	Purchase_Category	1.039428
8	Purchase_Amount	1.038464
9	Frequency_of_Purchase	1.018506
10	Purchase_Channel	1.020593
11	Brand_Loyalty	1.022461
12	Product_Rating	1.025410
13	Time_Spent_on_Product_Research(hours)	1.018591
14	Social_Media_Influence	1.124449
15	Discount_Sensitivity	1.021896
16	Return_Rate	1.021679
17	Customer_Satisfaction	1.024411
18	Engagement_with_Ads	1.140981
19	Device_Used_for_Shopping	1.222574
20	Payment_Method	1.022147
21	Customer_Loyalty_Program_Member	1.022796
22	Purchase_Intent	1.015391
23	Shipping_Preference	1.024747
24	Time_to_Decision	1.014817

Terlihat bahwa tidak terdapat variabel yang mengalami multikolinearitas, karena seluruh nilai VIF (Variance Inflation Factor) berada di bawah batas toleransi yaitu 10.

Result Interpretation

Modeling with probit regression

Berdasarkan hasil pemodelan regresi Probit, variabel Education_Level, Device_Used_for_Shopping, Payment_Method, Brand_Loyalty, dan Product_Rating berpengaruh terhadap penggunaan diskon karena memiliki p-value di bawah 0,05. Artinya, faktor-faktor seperti pendidikan, perangkat yang digunakan, cara pembayaran, loyalitas terhadap merek, dan penilaian produk ikut memengaruhi keputusan customer dalam menggunakan diskon. Metode pembayaran berpengaruh karena banyak promo yang diberikan lewat e-wallet atau kartu kredit tertentu. Customer yang setia pada suatu merek juga cenderung lebih tertarik menggunakan diskon dari merek tersebut. Sementara itu, produk dengan rating tinggi lebih dipercaya, sehingga membuat customer lebih yakin untuk memanfaatkan diskon saat membeli produk tersebut.



Interpretasi Hasil

Marginal Effects

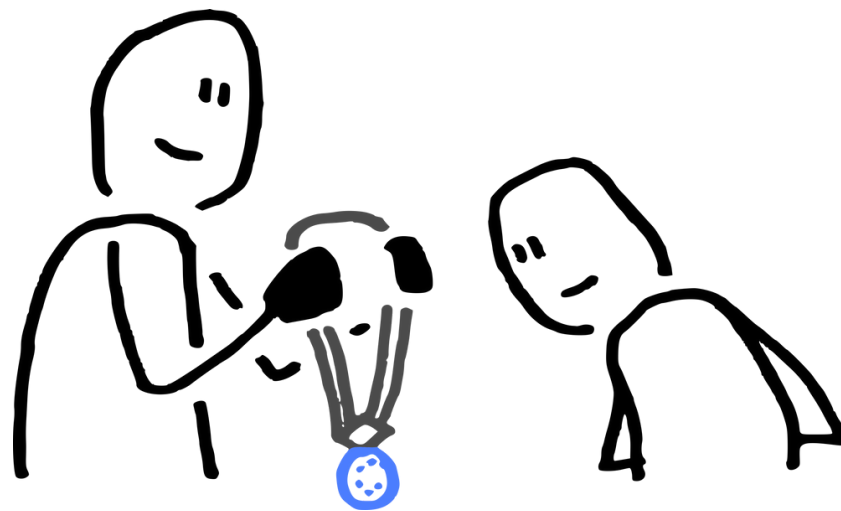
Berdasarkan grafik efek marginal, terlihat bahwa variabel dengan nilai dy/dx positif seperti **Payment_Method**, **Device_Used_for_Shopping**, dan **Education_Level** meningkatkan probabilitas customer untuk menggunakan diskon saat nilainya bertambah. Sebaliknya, variabel dengan nilai dy/dx negatif seperti **Brand_Loyalty** dan **Product_Rating** menurunkan probabilitas penggunaan diskon ketika nilainya meningkat. Dengan kata lain, semakin loyal customer terhadap merek dan semakin tinggi penilaian terhadap produk, maka semakin kecil probabilitas mereka memanfaatkan diskon. Sementara itu, faktor seperti metode pembayaran, perangkat yang digunakan untuk berbelanja, dan tingkat pendidikan justru meningkatkan kecenderungan customer untuk menggunakan diskon dalam transaksi.





*Politeknik Elektronika Negeri
Surabaya*

Thank you



The End