

CUSTOMER ANALYSIS

COMPREHENSIVE DATA ANALYTICS PROGRAM

› Start Slide

REFERENCE

DATASET SOURCE
[Kaggle Link](#)

COLAB NOTEBOOK
[Colab Link](#)



Background

The Company



ABC Company is a business engaged in the retail and distribution of consumer products, offering a wide range of categories including food, beverages (such as wine and alcohol), gold jewelry, as well as specialty products for children and teenagers.

Objective



The company aims to understand how customer profiles (such as age, income, family status, etc.) influence shopping habits and preferences for specific product categories.

The main objective is to enable marketing strategies to be targeted more effectively.



Business Understanding

SMART Framework

Specific

Analyze customer spending preferences across product categories.

Measurable

Measure total customer expenditure in each product category based on purchase transaction values and relate it to customer profile variables.

Achievable

Identify customer segments with similar preferences through a segmentation approach.

Relevant

Determine the most effective promotional strategies to maximize sales.

Time Bound

Develop new marketing strategies within one month.



Problem Statement

The objective of this project is to identify customer spending preferences across various product categories (food, wine, gold jewelry, children's and teenagers' products) based on customer characteristics such as age, marital status, number of children or teenagers in the household, and income level.

The results of this analysis will help the company develop personalized marketing strategies with more effective and targeted promotions.



Breaking Down the Problem

5W + 1H

How is the data structured, and what are the distribution patterns of customer characteristic variables (age, income, and family status)?

Is there a difference in the average spending on wine (MntWines) between married and unmarried customers?

Does customer age influence their preferred purchasing channel (web, catalog, physical store)?

Which products are most frequently purchased by customer segments based on age?

Through which channels do customers shop most often? (Website, catalog, physical store)

Are high-income customers more responsive to promotions?

Do newly joined customers (Dt_Customer) tend to be more active in purchasing products?

Data Understanding

The dataset consists of 2,240 rows and 28 columns. The variables in the dataset are grouped as follows:

- Year_Birth, Education, Marital_Status, Income:** Indicate customers' basic information and socio-economic status.
- Kidhome and Teenhome:** Represent the number of children and teenagers in the household.
- Recency and Dt_Customer:** Relate to the customer's most recent engagement.
- MntWines, MntFruits, MntMeatProducts, MntFishProducts, MntSweetProducts, MntGoldProds:** Show the amount spent on purchases in each product category.
- NumWebPurchases, NumCatalogPurchases, NumStorePurchases, NumWebVisitsMonth:** Describe customers' shopping habits across different channels.
- NumDealsPurchases, AcceptedCmp1, AcceptedCmp2, AcceptedCmp3, AcceptedCmp4, AcceptedCmp5, Response:** Indicate promotional activity and customer responses.

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

Data Cleaning

01

Removing Columns

Removed the columns Z_CostContact and Z_Revenue because they had no descriptions and were considered irrelevant.

02

Handling Missing Value

Identified missing values in the Income column (1.07%). These were handled by imputing the mean value.

03

Handling Data Inconsistency

Converted the Dt_Customer column data type from string to datetime format.

Index: 2240 entries, 5524 to 9405 Data columns (total 26 columns):			
#	Column	Non-Null Count	Dtype
0	Year_Birth	2240 non-null	int64
1	Education	2240 non-null	object
2	Marital_Status	2240 non-null	object
3	Income	2240 non-null	float64
4	Kidhome	2240 non-null	int64
5	Teenhome	2240 non-null	int64
6	Dt_Customer	2240 non-null	object
7	Recency	2240 non-null	int64
8	MntWines	2240 non-null	int64
9	MntFruits	2240 non-null	int64
10	MntMeatProducts	2240 non-null	int64
11	MntFishProducts	2240 non-null	int64
12	MntSweetProducts	2240 non-null	int64
13	MntGoldProds	2240 non-null	int64
14	NumDealsPurchases	2240 non-null	int64
15	NumWebPurchases	2240 non-null	int64
16	NumCatalogPurchases	2240 non-null	int64
17	NumStorePurchases	2240 non-null	int64
18	NumWebVisitsMonth	2240 non-null	int64
19	AcceptedCmp3	2240 non-null	int64
20	AcceptedCmp4	2240 non-null	int64
21	AcceptedCmp5	2240 non-null	int64
22	AcceptedCmp1	2240 non-null	int64
23	AcceptedCmp2	2240 non-null	int64
24	Complain	2240 non-null	int64
25	Response	2240 non-null	int64

dtypes: float64(1), int64(22), object(3)
memory usage: 472.5+ KB

What is the data structure and distribution pattern of customer characteristic variables (age, income, and family status)?

Statistic Descriptive

Analisis Statistik Usia Pelanggan

	Measure	Value
0	Mean	56.190000
1	Median	55.000000
2	Mode	49.000000
3	Standard Deviation	11.980000
4	Skewness	0.350000
5	Kurtosis	0.720000

Although the average customer income is quite high, most customers have incomes below the mean. This indicates a significant disparity in income distribution, with a small portion of customers having very high incomes.

The majority of customers are middle-aged, with an age concentration around **49–56 years old**. Given this relatively uniform distribution, marketing strategies can be focused on this age segment.

Analisis Statistik Pendapatan Pelanggan

	Measure	Value
0	Mean	52247.250000
1	Median	51741.500000
2	Mode	52247.251354
3	Standard Deviation	25037.800000
4	Skewness	6.800000
5	Kurtosis	161.390000

Statistic Descriptive

Distribusi Status Pernikahan Pelanggan

	Marital Status	Count
0	Married	864
1	Together	580
2	Single	480
3	Divorced	232
4	Widow	77
5	Alone	3
6	Absurd	2
7	YOLO	2

Mode: Married

Since most customers are married, marketing strategies targeting couples or families will be more relevant and potentially more effective.

Distribusi Tingkat Pendidikan Pelanggan

	Education Level	Count
0	Graduation	1127
1	PhD	486
2	Master	370
3	2n Cycle	203
4	Basic	54

Mode: Graduation

The majority of customers have a higher education background, allowing communication approaches to use more informative and rational language. Marketing strategies can be directed toward the upper-middle educated segment.

Statistic Descriptive

Jumlah Anak Kecil dalam Rumah Tangga (Kidhome)

Children (Kidhome)	Count
0	1293
1	899
2	48

Mode: 0

Most customers do not have young children at home, so child-themed products or promotions may be less relevant for the majority of customers. Marketing strategies should focus more on the needs of adult individuals or families without young children.

Jumlah Anak Remaja dalam Rumah Tangga (Teenhome)

Teenagers (Teenhome)	Count
0	1158
1	1030
2	52

Mode: 0

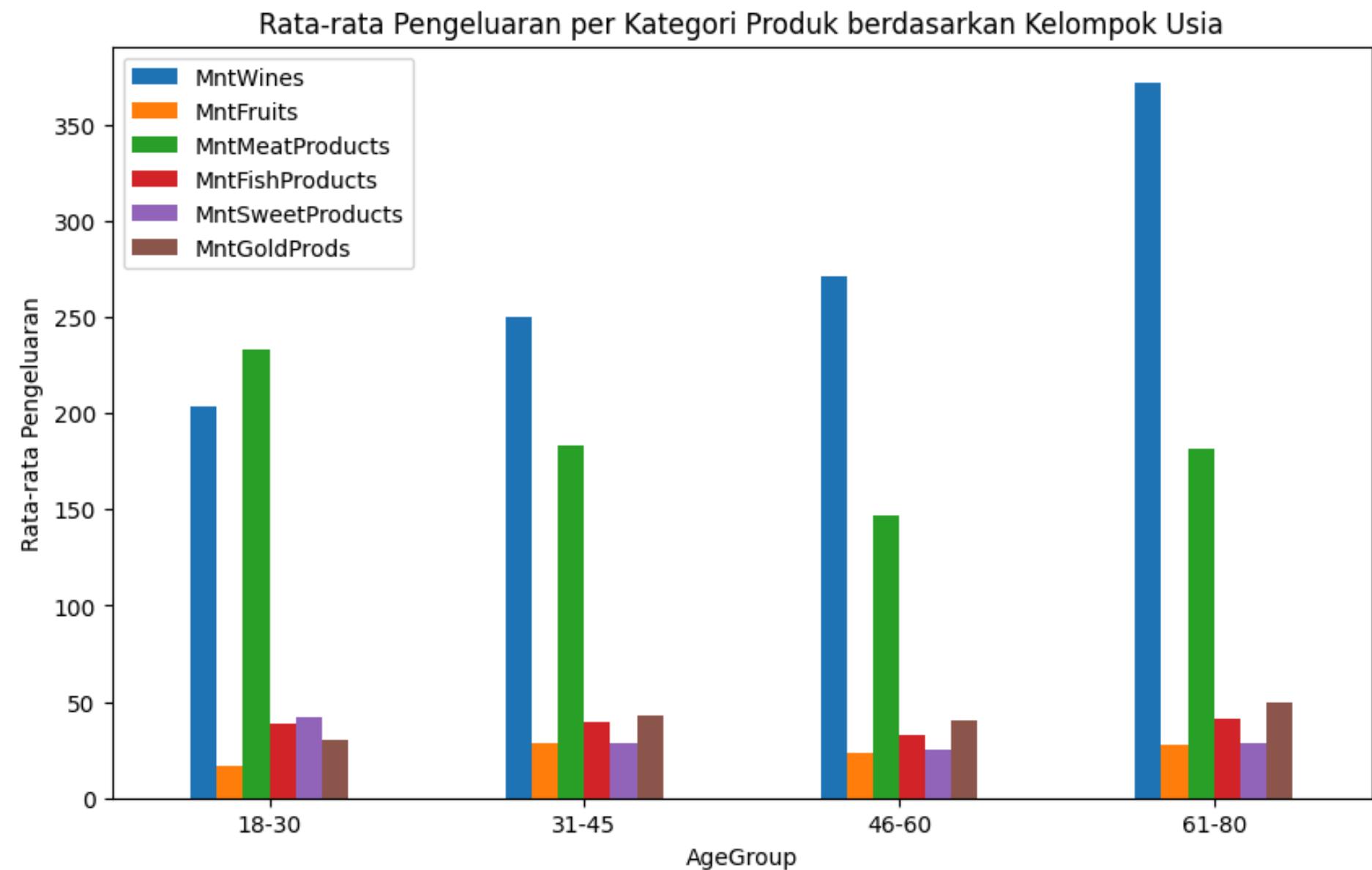
The majority of customers also do not live with teenagers, although the number who have one teenager at home is still significant (1,030 customers). This means that while families with teenagers are not the dominant segment, there is still potential to target them as a market.



Which products are most frequently purchased by customer segments based on age?

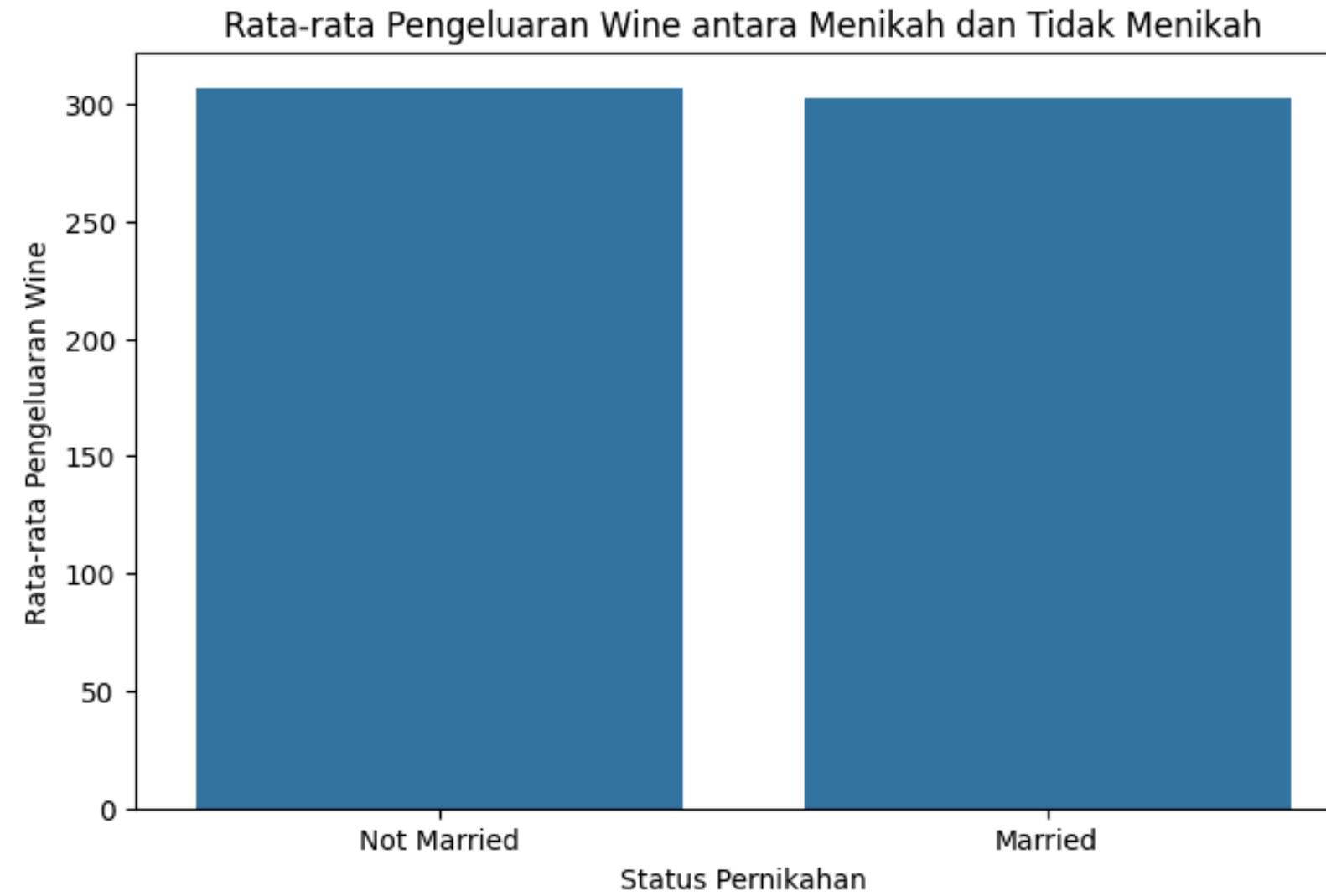
Products with the Highest Interest

- Older consumers (**ages 61–80**) tend to spend more on wine.
- The **18–45 age group** dominates spending on meat products.
- Spending patterns for **sweet products** and **gold jewelry** are relatively **consistent across age groups**.
- **Fruit and fish** products account for a **smaller share of total spending**.



Is there a difference in the average spending on wine (MntWines) between married and unmarried customers?

Differences in Wine Consumption



Hypotheses

- **Null Hypothesis (H_0):** There is no significant difference in the average spending on wine between married and unmarried customers.
- **Alternative Hypothesis (H_1):** There is a significant difference in the average spending on wine between married and unmarried customers.

Significance Level

Alpha = 5%

Test Statistics

Based on the results of the independent t-test:

- T-statistic: -0.28
- P-value: 0.7758

Decision

Since the p-value > 0.05, we fail to reject the null hypothesis (H_0). In other words, the null hypothesis is accepted.

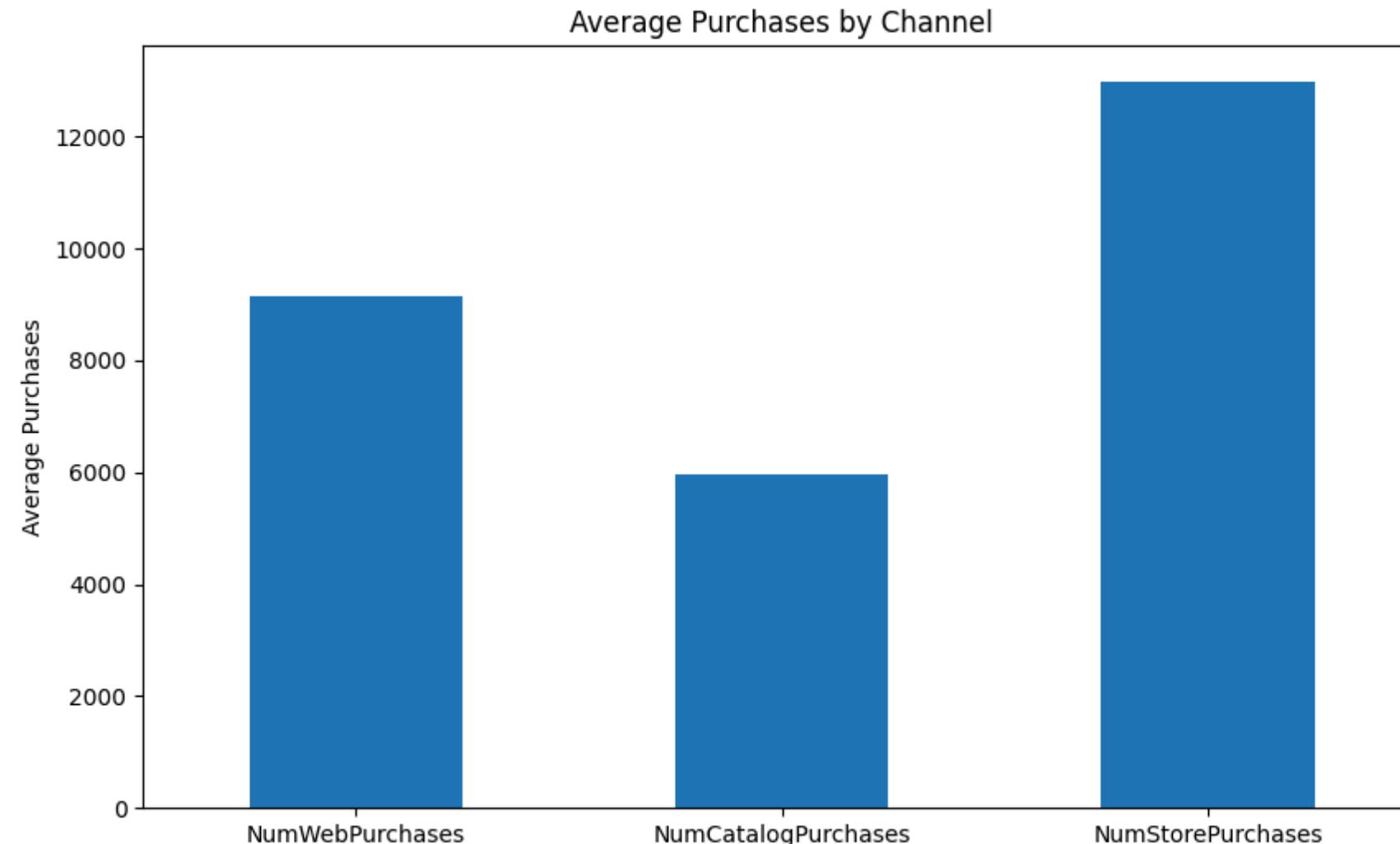
Conclusion

This means there is no significant difference in the average wine spending between married and unmarried customers.



Through which channels do customers shop most often (website, catalog, physical store)?

Highest Number of Purchases



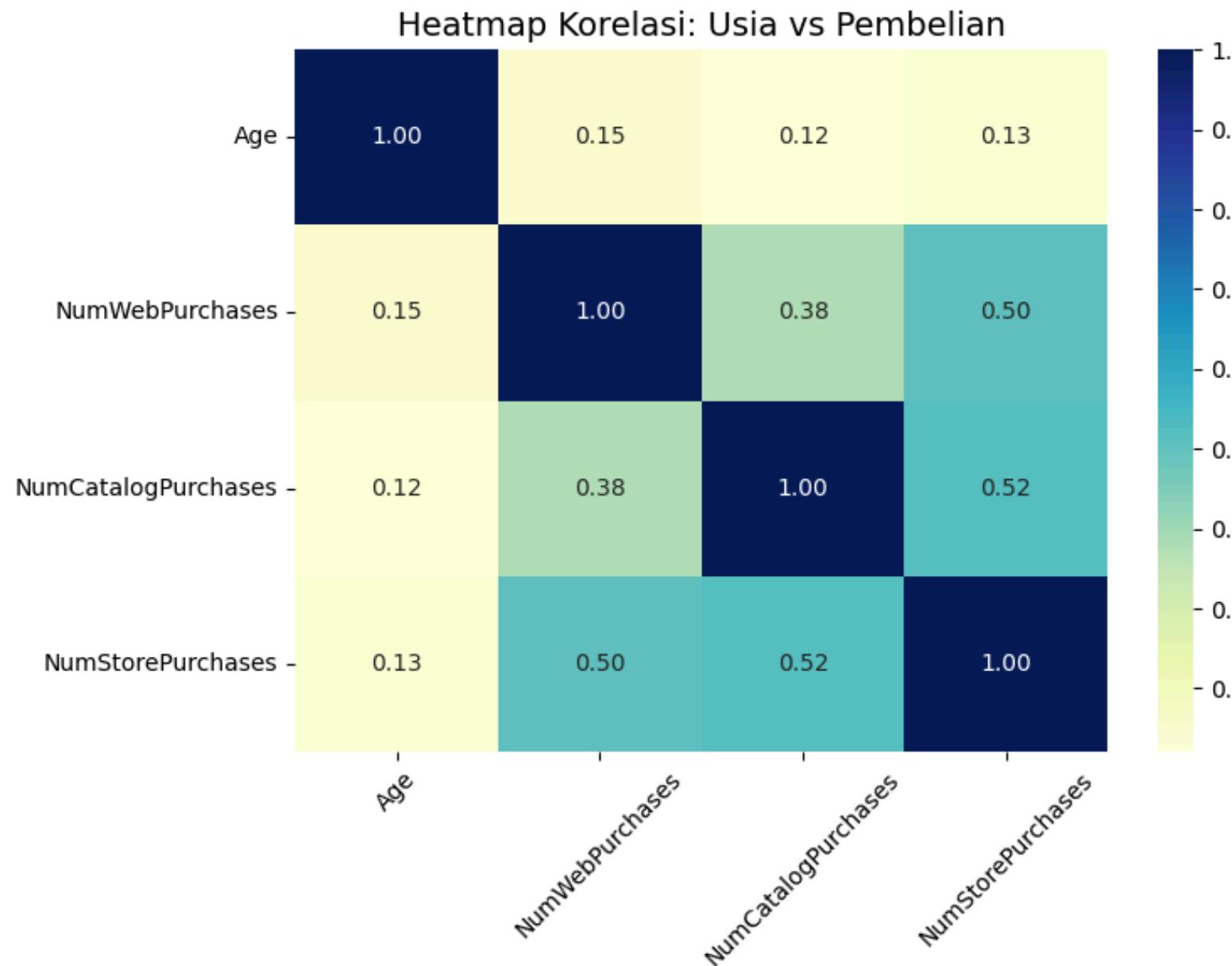
Purchases are predominantly made through physical stores (NumStorePurchases), followed by purchases via the website (NumWebPurchases), and lastly through catalogs (NumCatalogPurchases). Therefore:

- **Marketing strategies can be focused on physical stores** by adding services that enhance the customer experience, such as loyalty programs or exclusive events.
- **The website can be improved** by increasing user convenience, speeding up the checkout process, and offering discount vouchers.
- **It is necessary to evaluate the effectiveness of the catalog channel.** If it is found to be inefficient, consideration should be given to replacing it with other digital media.



Does customer age influence their preferred purchasing channel (website, catalog, physical store)?

Correlation Between Age and Channels



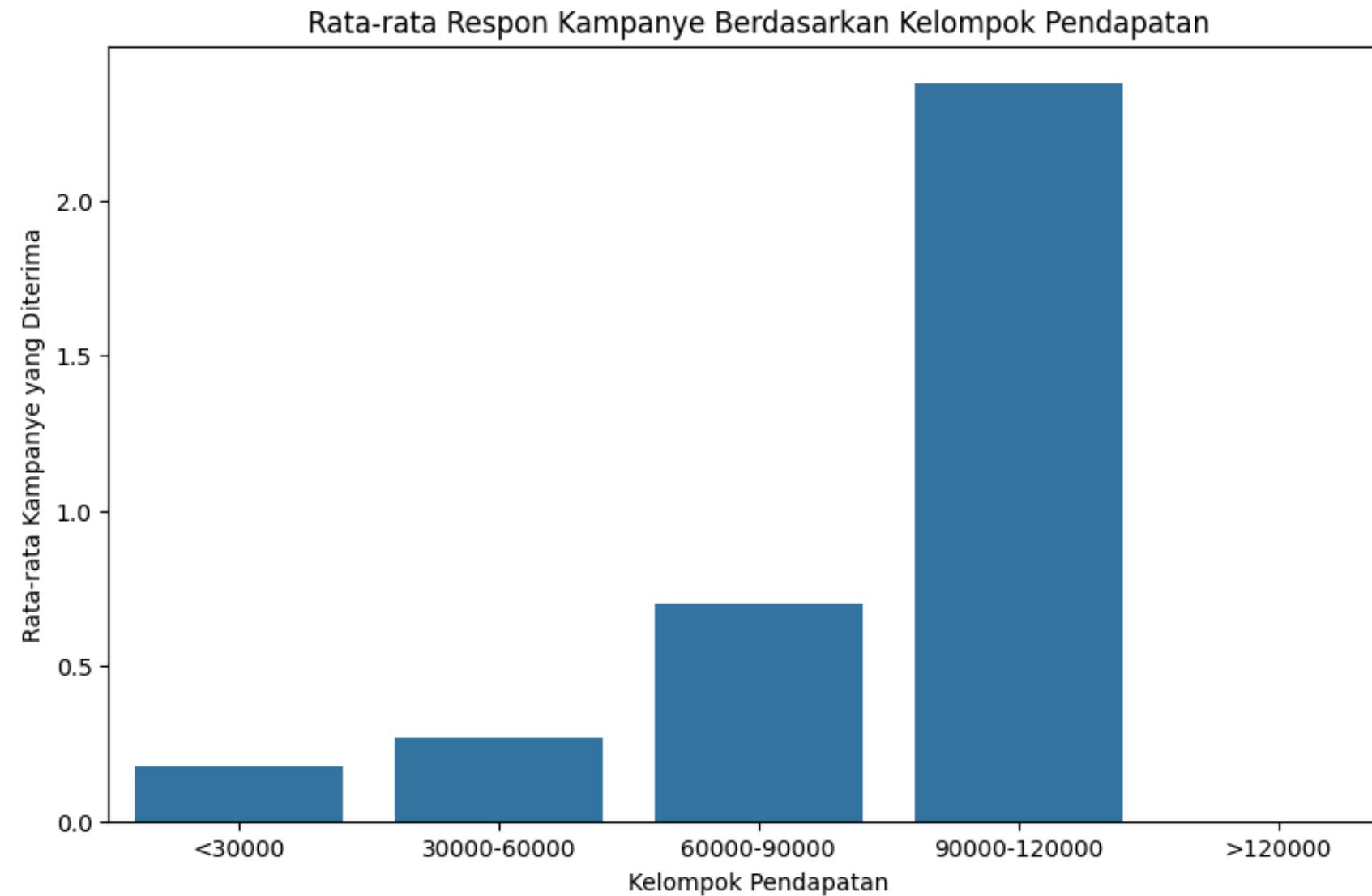
The correlation between age and the different purchasing channels is relatively low:

- **Age vs. NumWebPurchases: 0.15**
- **Age vs. NumCatalogPurchases: 0.12**
- **Age vs. NumStorePurchases: 0.13**

This indicates that age does not strongly influence customers' habits in choosing purchasing channels. Customers across different age groups have fairly diverse preferences. Because the correlations are low, channel-based strategies do not need to be heavily focused on age segmentation.

**Are high-income customers more
responsive to promotions?**

Campaign Response by Income Group



Campaign Response by Income Group

- **Income 90,000–120,000:** The highest response rate, but a small number of customers (45 people). Therefore, a personalized, small-scale approach can be continued for this group.
- **Income 30,000–60,000:** The largest segment (1,029 customers), but with a low response rate. This group can be optimized through broad, large-scale promotions.
- **Income >120,000:** High response, but only 8 customers, which is not representative. Therefore, major decisions based on this group should be avoided.

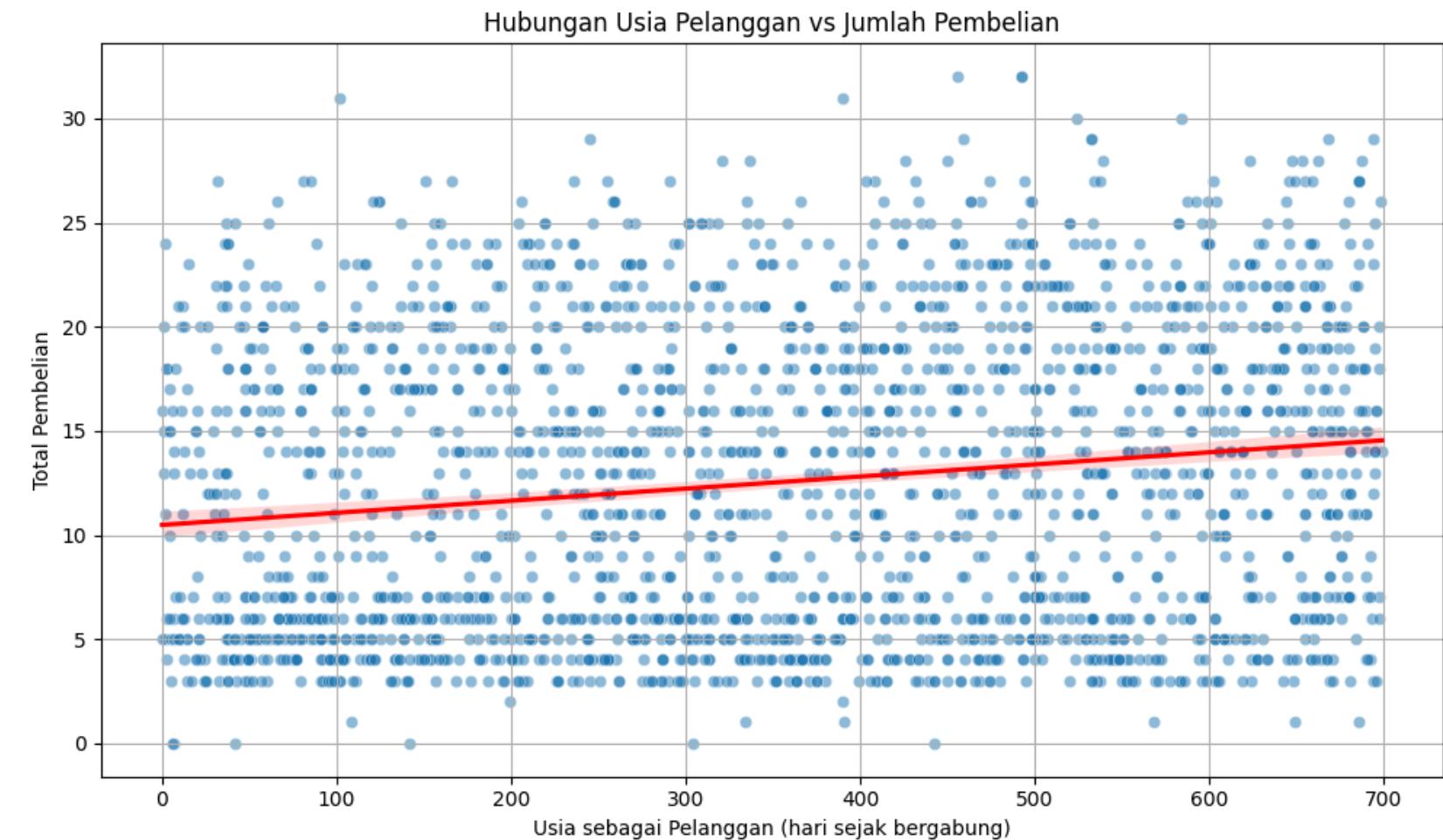


**Do newly joined customers (Dt_Customer)
tend to be more active in purchasing
products?**

Product Purchase Tendencies

A correlation value of 0.163 between customer tenure (calculated from Dt_Customer) and the total number of purchases (the sum of NumWebPurchases, NumCatalogPurchases, and NumStorePurchases) indicates a weak positive relationship.

This means that the longer someone has been a customer, the more purchases they tend to make; however, the effect is very limited.



Conclusion & Recomendation

- The products with the **highest spending** are **wine** (among older customers) and **meat** (among younger customers).
- No significant differences were found in wine consumption based on marital status, and age does not strongly influence channel preferences.
- Most purchases are made in **physical stores**, followed by the website.
- **High-income customers** tend to be **more responsive** to promotions, although their numbers are limited.
- Customer tenure shows a slightly positive but weak correlation with the number of purchases.



Conclusion & Recomendation

- Focus strategies on the **49–56 age segment** with an educational approach.
- Optimize **wine promotions for older customers** and **meat promotions for younger customers**.
- Prioritize the physical store channel and enhance the shopping experience.
- Offer **general promotions for middle-income customers** and adopt a **personalized approach for the premium segment**.
- Evaluate the effectiveness of the catalog channel and consider shifting to digital media.



THANK YOU

COMPREHENSIVE DATA ANALYTICS PROGRAM

› End Slide