



CUSTOMER SEGMENTATION

Giới thiệu bài toán

Phân tích và phân loại khách hàng giúp doanh nghiệp **hiểu biết sâu sắc hơn về khách hàng**, cho phép điều chỉnh sản phẩm để đáp ứng nhu cầu, hành vi và mối quan tâm riêng biệt của các loại khách hàng khác nhau.

Thay vì phân bổ nguồn lực để tiếp thị một sản phẩm mới cho toàn bộ cơ sở dữ liệu khách hàng, các công ty có thể **xác định các phân khúc** có nhiều khả năng quan tâm đến sản phẩm nhất. Sau đó, các nỗ lực **tiếp thị** có thể được **hướng tới các phân khúc cụ thể, tối ưu hóa** việc sử dụng **tài nguyên**.

Mục đích phân tích

- Tìm hiểu **hành vi khách hàng**
- Xác định **nhóm khách hàng**
- **Đề xuất chiến lược** để thu hút khách hàng mới và giữ chân khách hàng cũ nhằm tối đa doanh thu và giảm thiểu chi phí



MỤC LỤC

- 01 • **TIỀN XỬ LÝ DỮ LIỆU**
- 02 • **PHÂN TÍCH HÀNH VI KHÁCH HÀNG**
- 03 • **XÂY DỰNG MÔ HÌNH**
- 04 • **ĐỀ XUẤT CHIẾN LƯỢC**

TIỀN XỬ LÝ DỮ LIỆU

Bộ dữ liệu gồm **29 cột** và **2240 dòng**

```
df.info()  
  
<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)
```

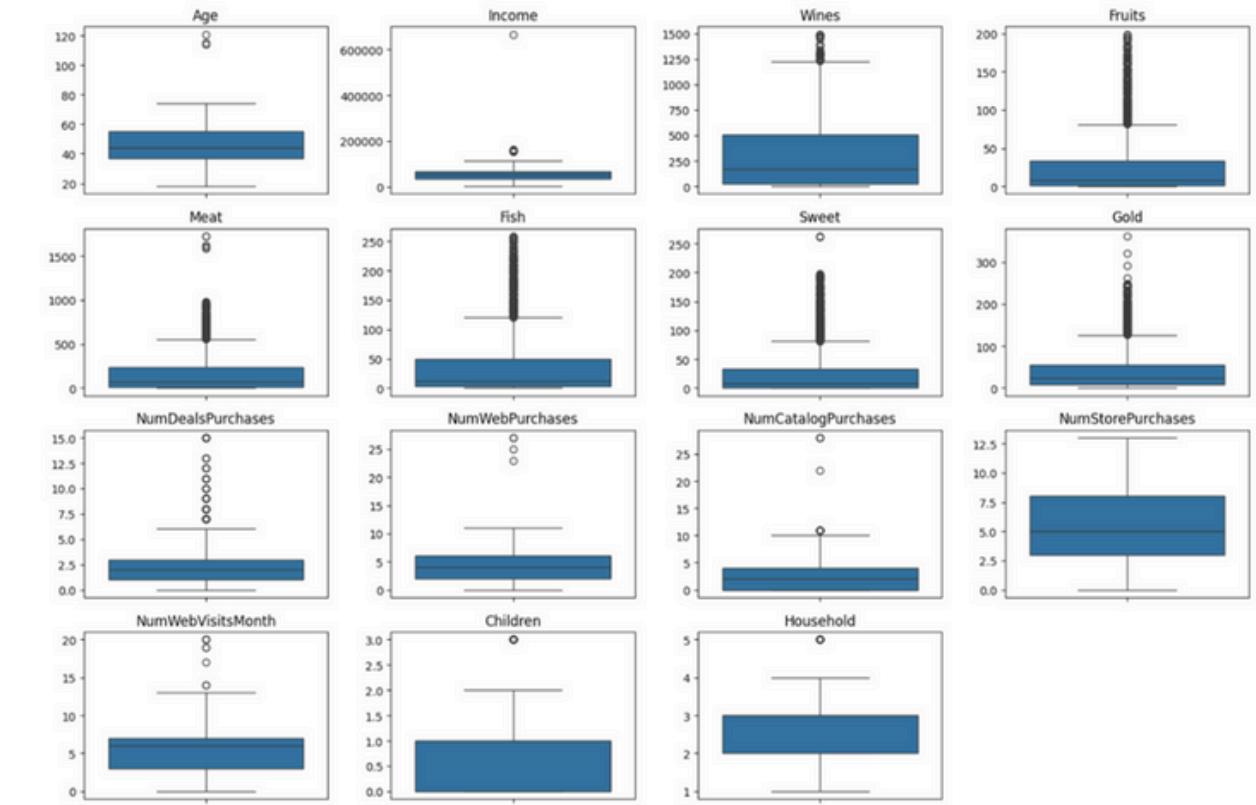
Kiểm tra dữ liệu null

```
df.isnull().sum()  
  
ID           0  
Year_Birth   0  
Education    0  
Marital_Status 0  
Income        24  
Kidhome     0  
Teenhome    0
```

Sử dụng **số trung vị** (median) để xử lý **24** dữ liệu null ở cột **Income**

```
df['Income'].fillna(df['Income'].median(), inplace=True)
```

Kiểm tra Outliner



Bộ dữ liệu sau khi xử lý Outliner bằng IQR

```
<class 'pandas.core.frame.DataFrame'>  
Index: 964 entries, 1 to 2239
```

TIỀN XỬ LÝ DỮ LIỆU

Tạo cột Children

Kết hợp 02 cột **Kidhome** và **Teenhome**

```
df2['Children']=df['Kidhome'] + df['Teenhome']
```

Bỏ 02 cột **Kidhome** và **Teenhome**

```
df2 = df2.drop(['Kidhome', 'Teenhome'], axis = 1)
```

Tạo cột Household

Tạo cột mới là **People**, nếu Married là trả về 2 người, còn Single sẽ trả về 1 người. Chuyển kiểu dữ liệu từ chuỗi sang int

```
df2['People']=df2['Marital_Status'].replace(["Married", "Single"], ['2','1'])  
People=df2['People'].astype("int64")
```

Tạo cột **Household** trả về số lượng thành viên trong một hộ gia đình

```
df2['Household']=df2['Children']+People
```

Tạo, biến đổi thêm các cột mới

Tạo cột Age từ cột Year_Birth

```
df['Age'] = 2014 - df['Year_Birth']
```

Biến đổi dữ liệu cột Education

```
df2['Education'] = df2['Education'].replace({"Basic": "Undergraduate",  
"2n Cycle": "Undergraduate",  
"Graduation": "Graduate",  
"Master": "Postgraduate",  
"PhD": "Postgraduate"})
```

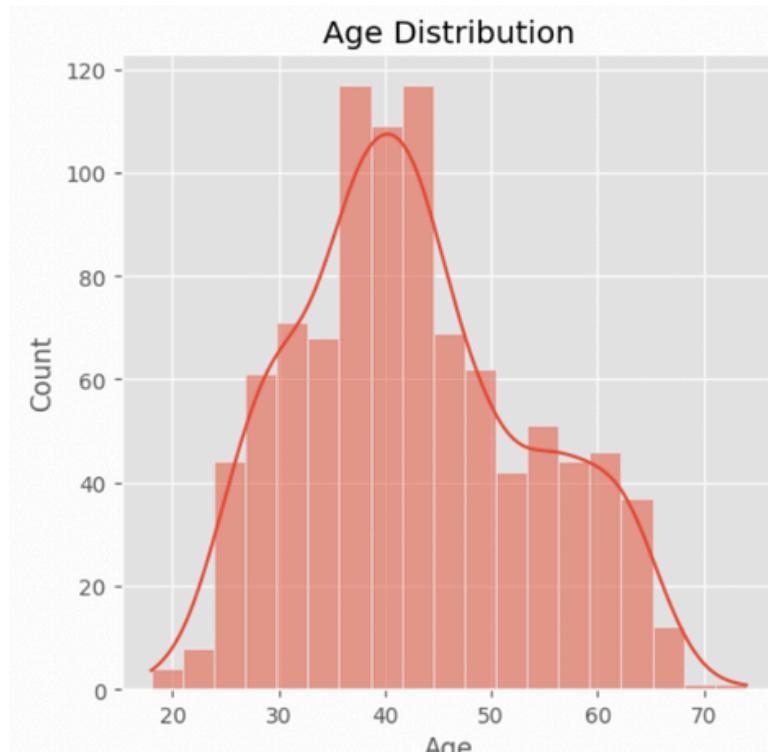
Biến đổi dữ liệu cột Marital_Status

```
df2['Marital_Status'] = df2['Marital_Status'].replace({"Together": "Married",  
"Divorced": "Single",  
"Widow": "Single",  
"Alone": "Single",  
"Absurd": "Single",  
"YOLO": "Single"})
```

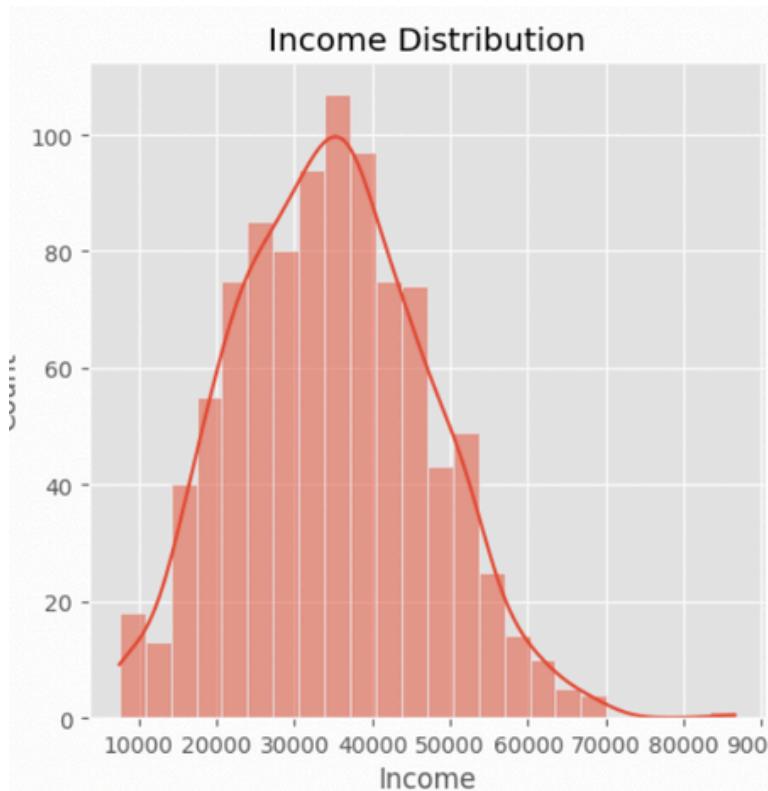
PHÂN TÍCH HÀNH VI KHÁCH HÀNG

Nhân khẩu học

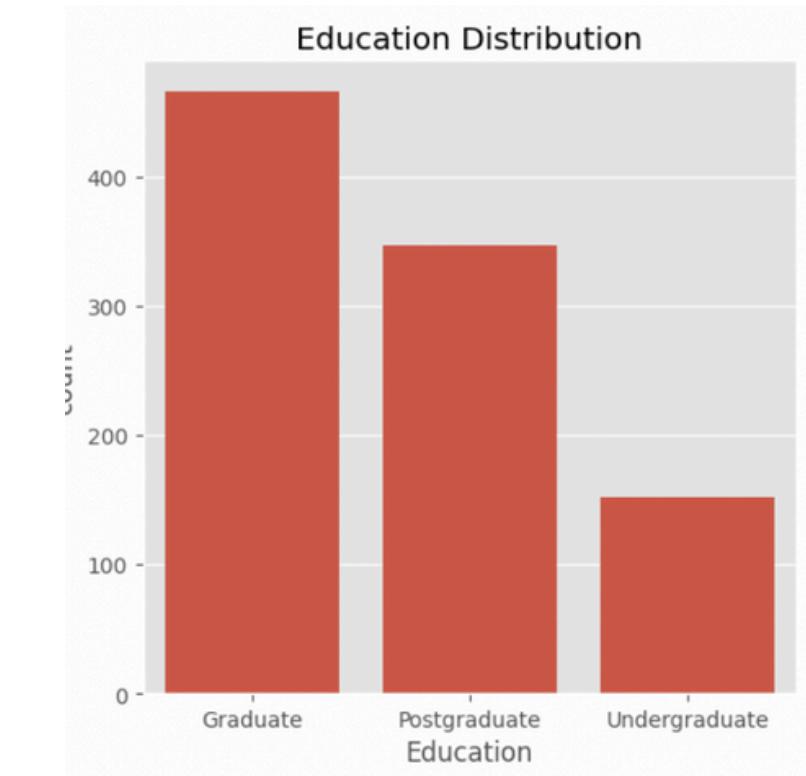
Độ tuổi



Thu nhập



Trình độ học vấn



Đa số khách hàng thuộc nhóm tuổi từ **30 đến 50**, nhiều nhất là trong khoảng **35-40** tuổi. Đây là độ tuổi chính mà doanh nghiệp có thể tập trung tiếp cận.

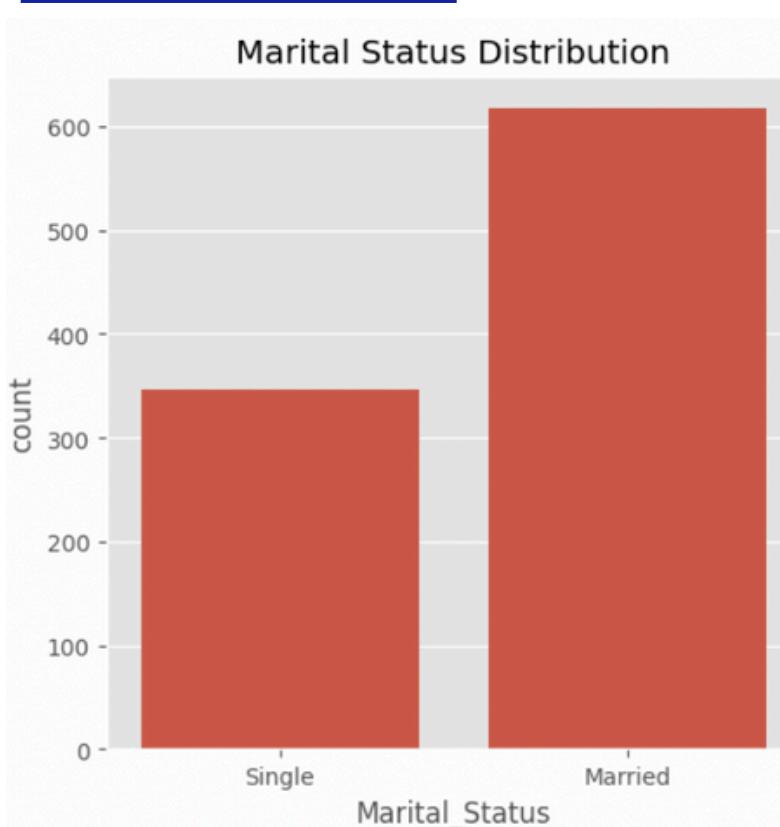
Thu nhập của khách hàng phân bố theo hình dạng chuông, với đa số trong khoảng **30,000 đến 40,000**, cho thấy họ có **khả năng chi tiêu ở mức trung bình khá**.

Số lượng người có **bằng Đại học** chiếm tỷ lệ cao nhất, tiếp theo là người có bằng Sau đại học. Đa số khách hàng có học vấn cao nên có thể họ **hiểu biết và có yêu cầu về chất lượng dịch vụ/sản phẩm**.

PHÂN TÍCH HÀNH VI KHÁCH HÀNG

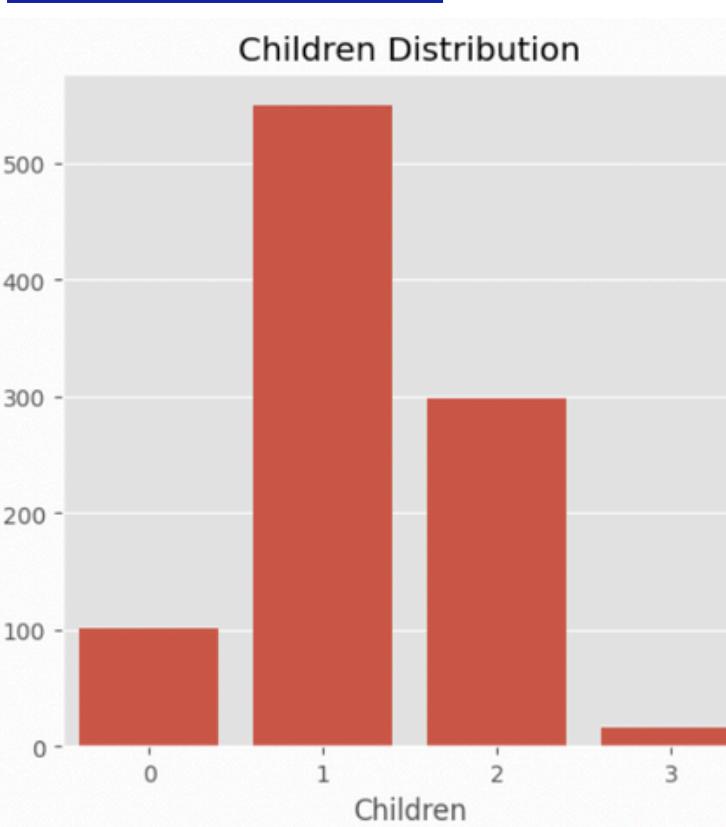
Nhân khẩu học

Tình trạng hôn nhân



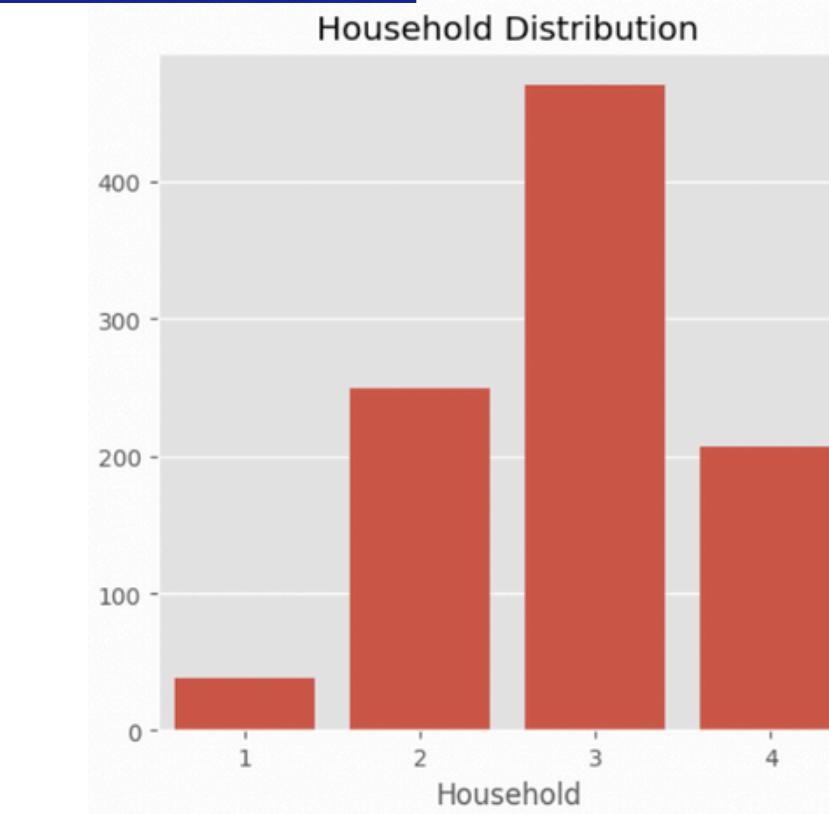
Số lượng **khách hàng đã kết hôn** nhiều hơn hẳn so với số lượng khách hàng độc thân, có thể phần lớn **quan tâm** nhiều hơn đến các **sản phẩm liên quan đến gia đình**.

Số lượng con cái



Đa số khách hàng có **1 con**, tiếp theo là không có con và có 2 con. Số lượng khách hàng có 3 con rất ít.

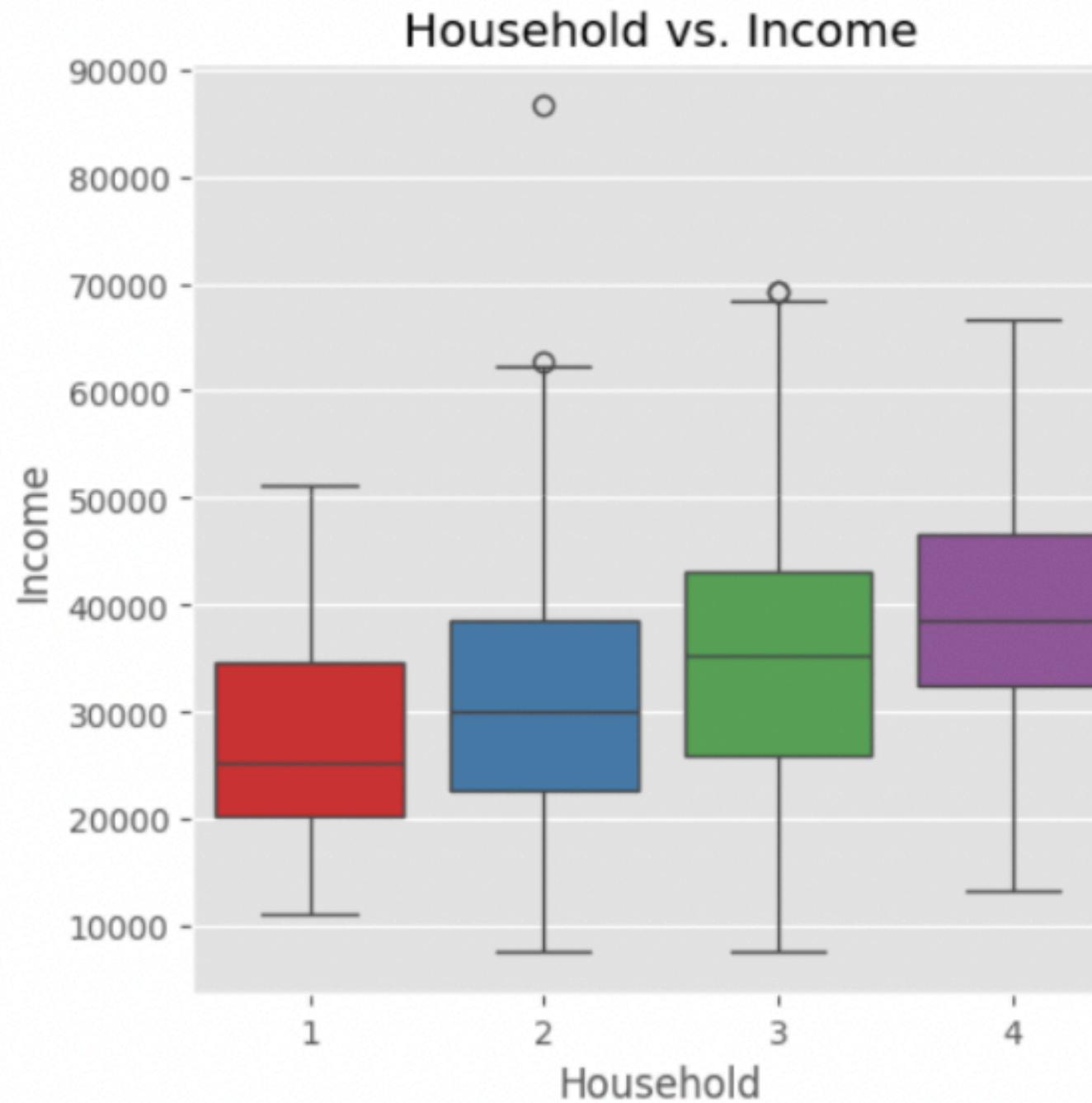
Số thành viên trong hộ gia đình



Hầu hết các hộ gia đình có **3 thành viên**, tiếp theo là hộ gia đình có 2 thành viên và 4 thành viên. Điều này cho thấy các **hộ gia đình nhỏ** là **đối tượng khách hàng chủ yếu**.

PHÂN TÍCH HÀNH VI KHÁCH HÀNG

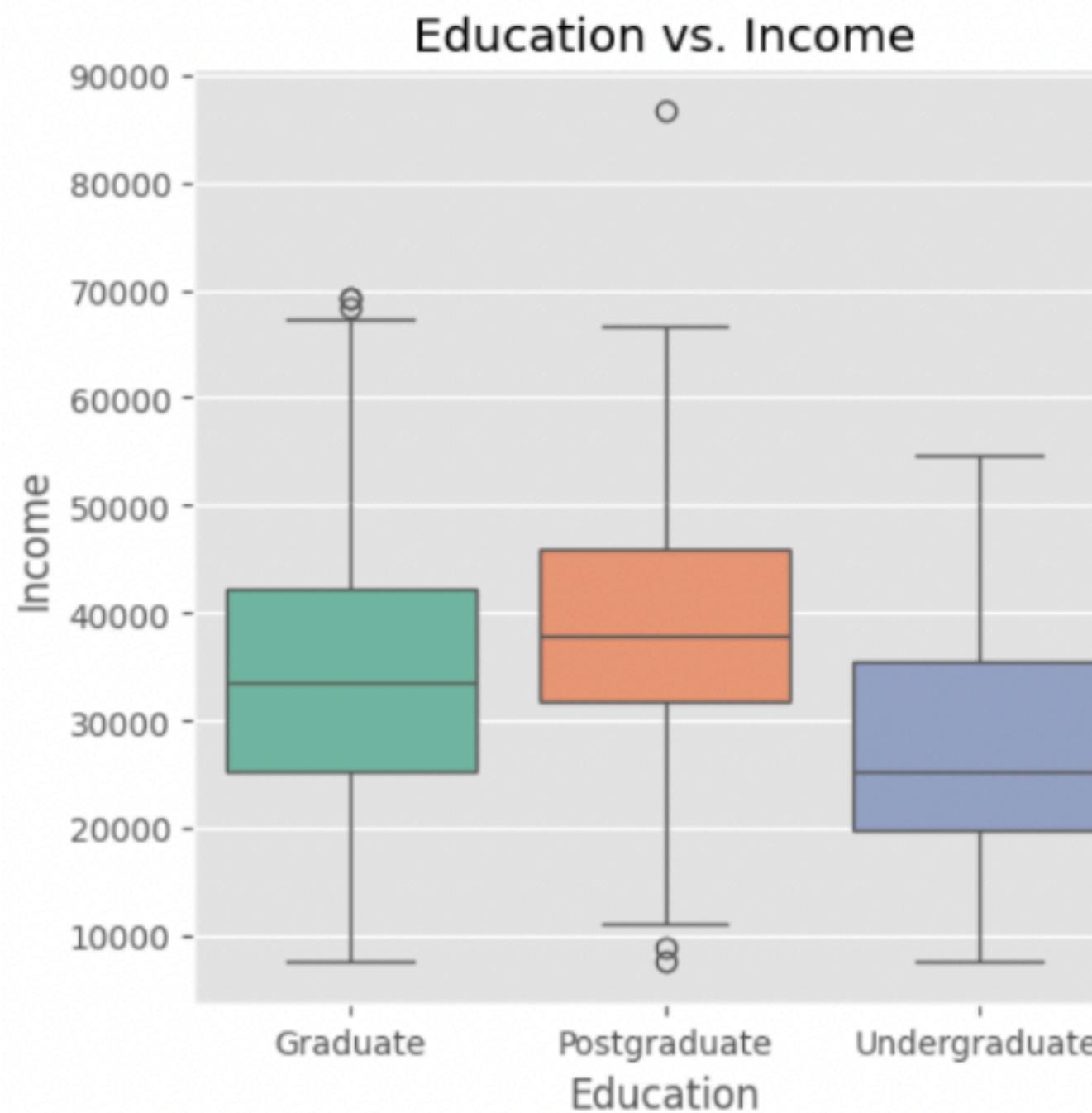
Mối quan hệ giữa **Số thành viên trong hộ gia đình** và **Thu nhập**



- Hộ gia đình có **số thành viên lớn** (3 hoặc 4 thành viên) thường có **thu nhập cao hơn** so với những hộ gia đình nhỏ hơn (1 hoặc 2 thành viên). Điều này có thể cho thấy các hộ gia đình lớn hơn có nhu cầu tiêu dùng cao hơn và có khả năng chi trả tốt hơn.
- Hộ gia đình có **4 thành viên có mức thu nhập trung bình cao nhất**, với mức phân bố tương đối đồng đều. Trong khi đó, hộ gia đình **1 thành viên** có **thu nhập trung bình thấp nhất**.
- Có một số giá trị ngoại lệ (outliers) trong cả 4 nhóm, cho thấy có những hộ gia đình có thu nhập vượt trội hoặc thấp hơn đáng kể so với trung bình.

PHÂN TÍCH HÀNH VI KHÁCH HÀNG

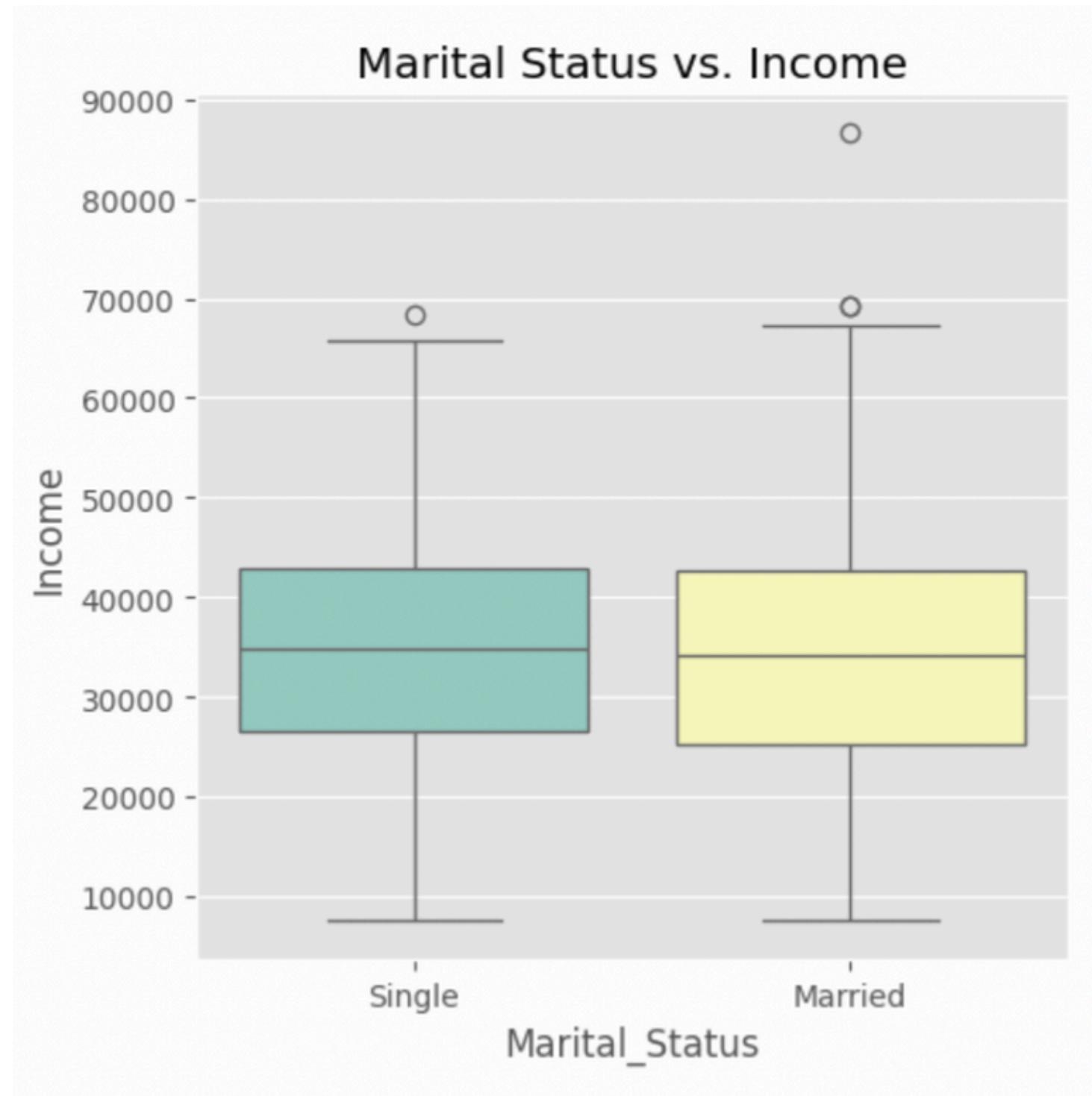
Mối quan hệ giữa **Trình độ học vấn** và **Thu nhập**



- Những người có **bằng Sau đại học** (Postgraduate) có **thu nhập trung bình cao hơn** so với những người có bằng Đại học (Graduate) và bằng cấp thấp hơn (Undergraduate). Điều này cho thấy **trình độ học vấn cao hơn có liên quan đến mức thu nhập cao hơn**.
- Tuy nhiên, cũng có một vài **trường hợp ngoại lệ** trong cả ba nhóm, với một số người có trình độ học vấn thấp hơn nhưng thu nhập lại cao hơn nhiều so với trung bình.

PHÂN TÍCH HÀNH VI KHÁCH HÀNG

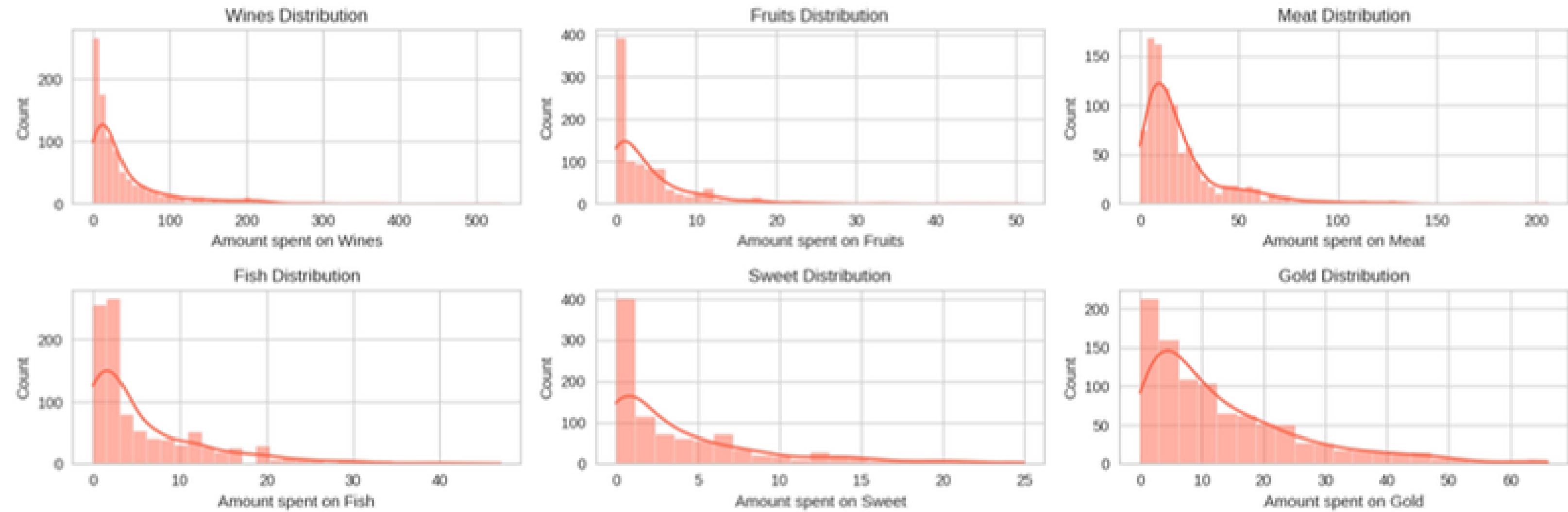
Mối quan hệ giữa **Tình trạng hôn nhân** và **Thu nhập**



- Thu nhập của những **người đã kết hôn** (Married) và những **người độc thân** (Single) có mức **trung bình tương đương nhau**, không có sự chênh lệch lớn về thu nhập giữa hai nhóm này. **Tình trạng hôn nhân không ảnh hưởng quá lớn đến thu nhập**, cho thấy rằng các chiến lược tiếp thị có thể không cần phải phân biệt quá nhiều giữa khách hàng đã kết hôn và độc thân trong khía cạnh thu nhập.
- Cả hai nhóm đều có sự **phân bố thu nhập rộng**, với một số giá trị ngoại lệ cho thấy có những cá nhân có thu nhập rất cao so với số đông

PHÂN TÍCH HÀNH VI KHÁCH HÀNG

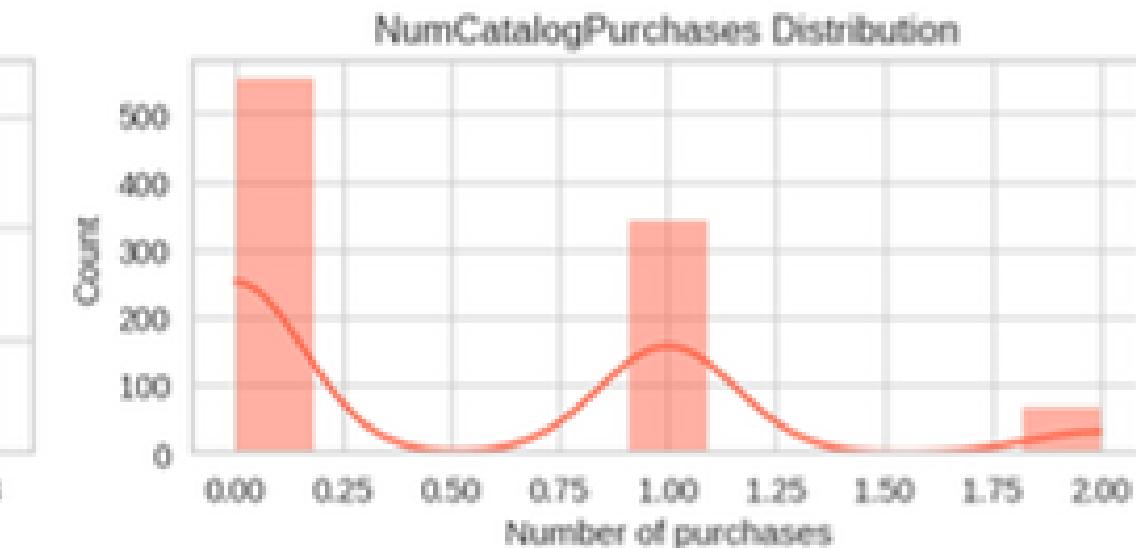
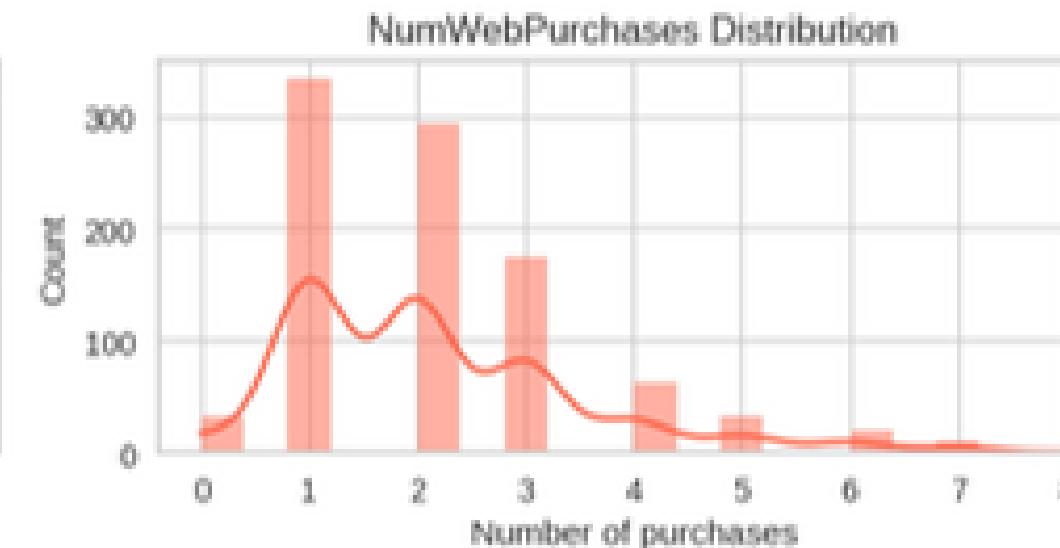
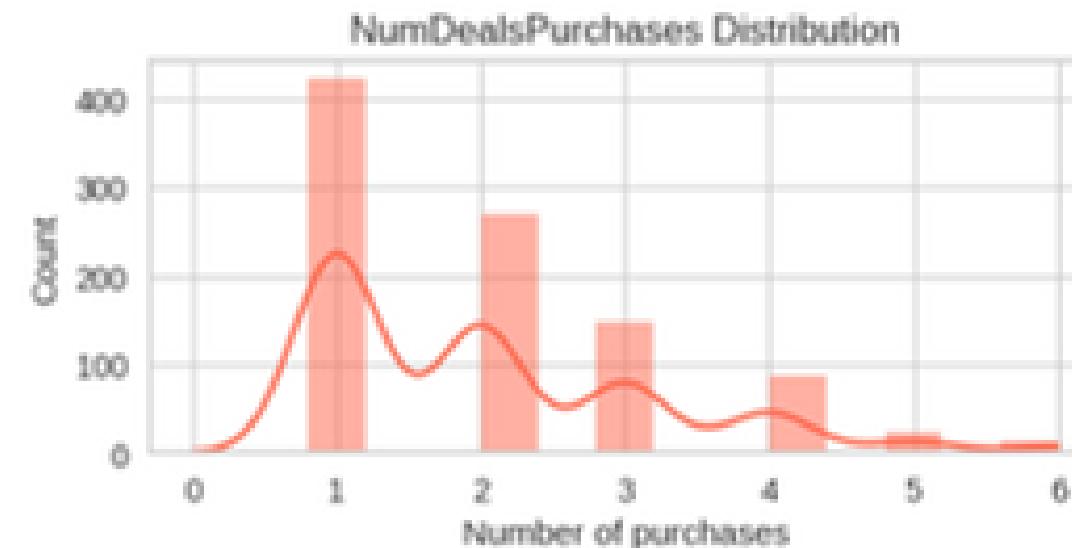
Hành vi mua hàng



- Đối với **tất cả các danh mục sản phẩm**, hầu hết khách hàng có xu hướng **mua với số lượng nhỏ**. Lượng tiền chi cho sản phẩm **càng nhiều**, lượng đơn hàng **càng giảm**. Điều này cho thấy phần lớn khách hàng có xu hướng mua sắm một cách tiết kiệm hoặc chỉ mua những gì họ cần trong ngắn hạn.
- Trái cây** (Fruits) và **Đồ ngọt** (Sweet) là mặt hàng được mua lẻ nhiều nhất. Các mặt hàng khác như **Rượu** (Wines), **Thịt** (Meat), **Cá** (Fish) và **Vàng** (Gold) có số lượng mua lẻ thấp hơn đáng kể.

PHÂN TÍCH HÀNH VI KHÁCH HÀNG

Hành vi mua hàng



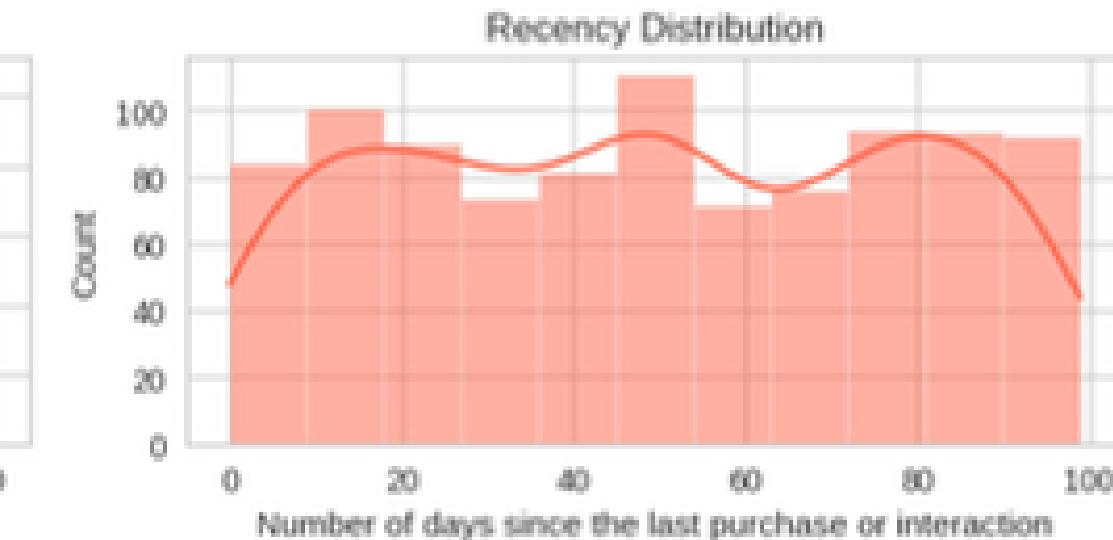
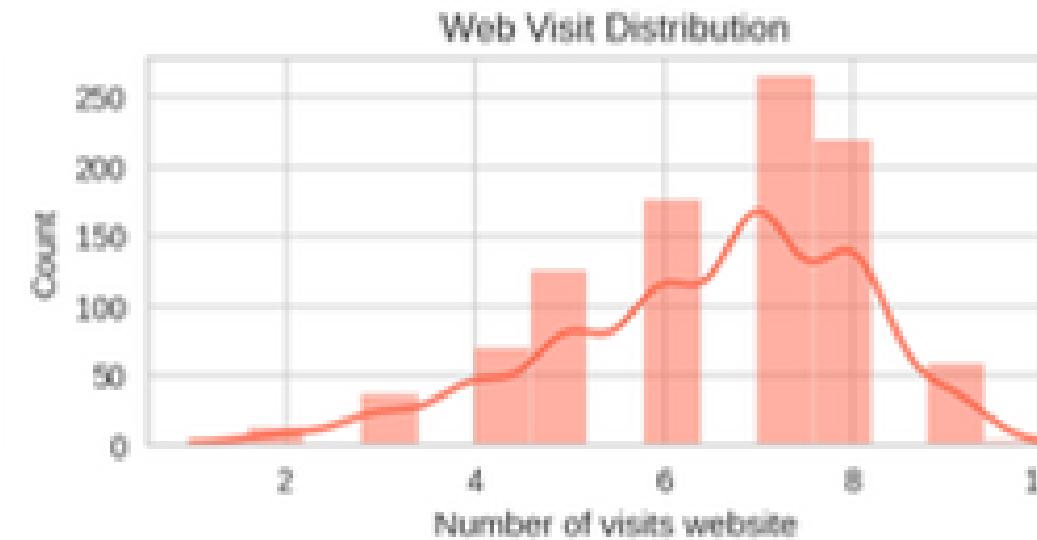
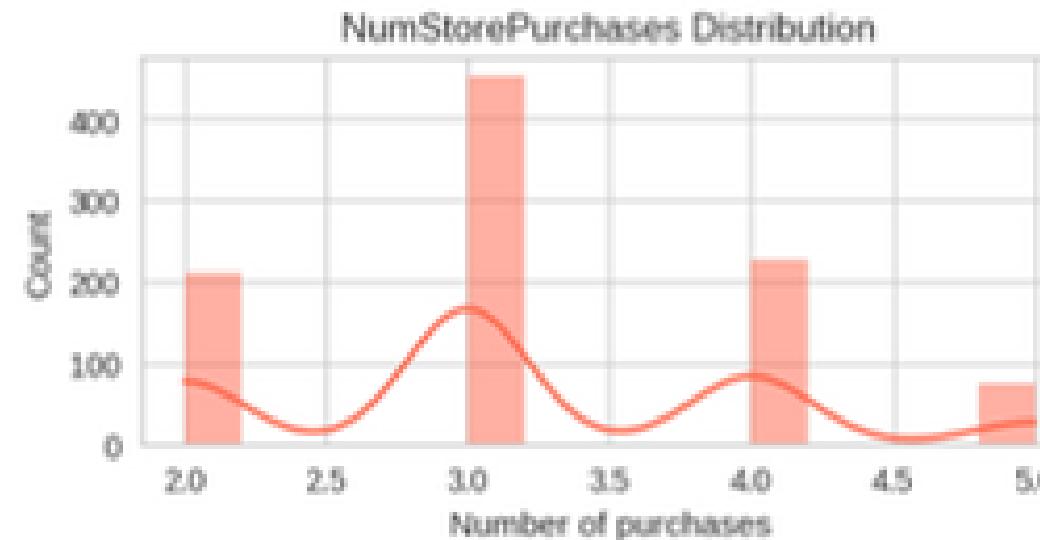
Phần lớn khách hàng tham gia vào các giao dịch mua hàng theo chương trình khuyến mãi **chỉ một vài lần**. Điều này cho thấy **chương trình khuyến mãi có thể không phải là yếu tố chính thu hút khách hàng**, hoặc có thể là do số lượng chương trình khuyến mãi có giới hạn.

Số lượng giao dịch mua hàng **qua web** cũng tập trung chủ yếu ở mức **thấp**, với phần lớn khách hàng thực hiện từ **1 đến 3 giao dịch**. Điều này có thể cho thấy rằng **trải nghiệm mua sắm qua web chưa được** khách hàng **ưu tiên** hoặc có thể họ thích mua trực tiếp tại cửa hàng hơn.

Giao dịch mua qua các ấn phẩm quảng cáo (catalog) **cực ít**. Chúng tỏ các khách hàng thường **không đọc** các ấn phẩm quảng cáo.

PHÂN TÍCH HÀNH VI KHÁCH HÀNG

Hành vi mua hàng



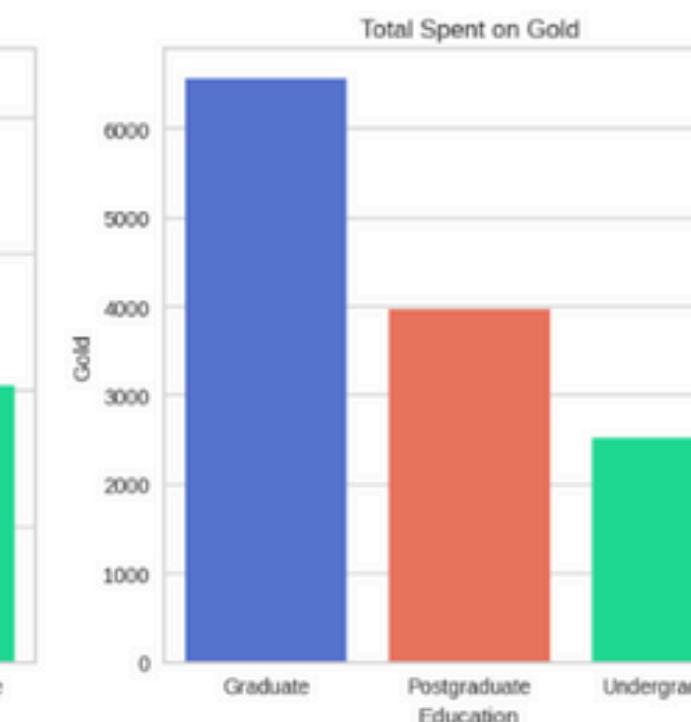
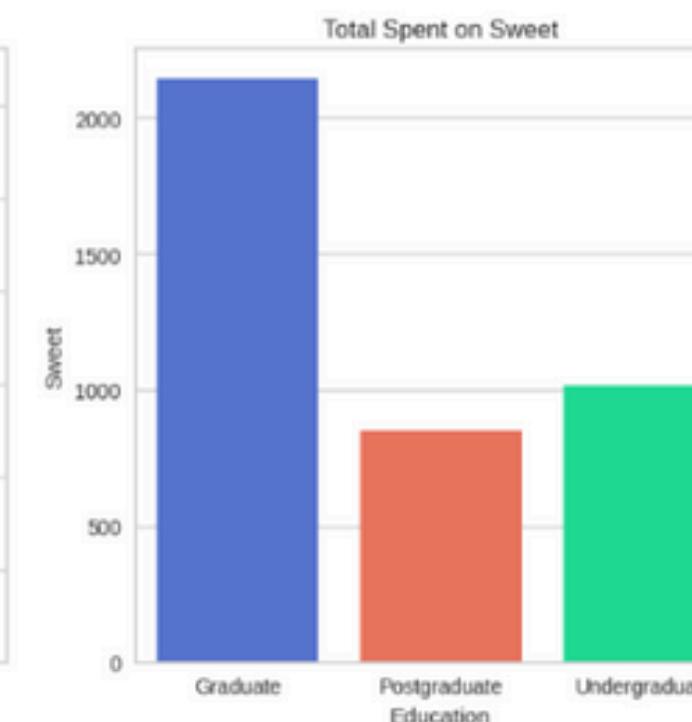
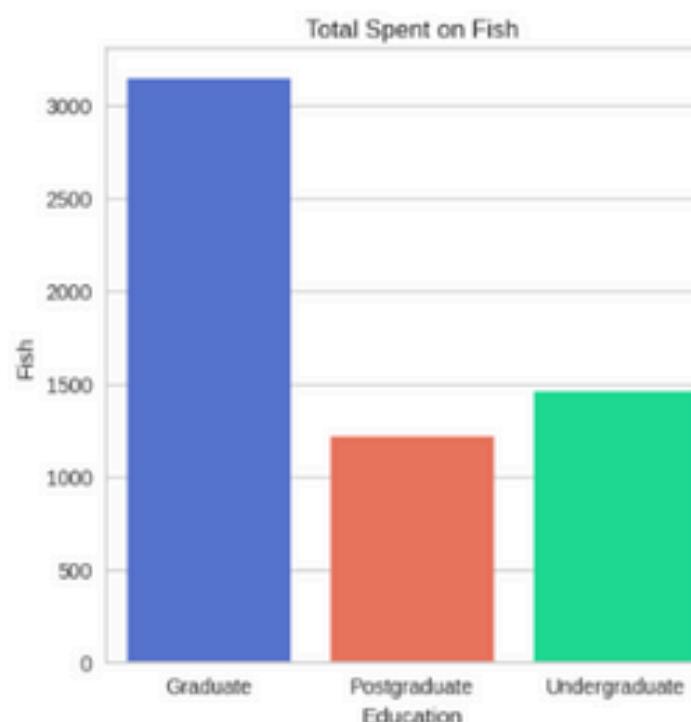
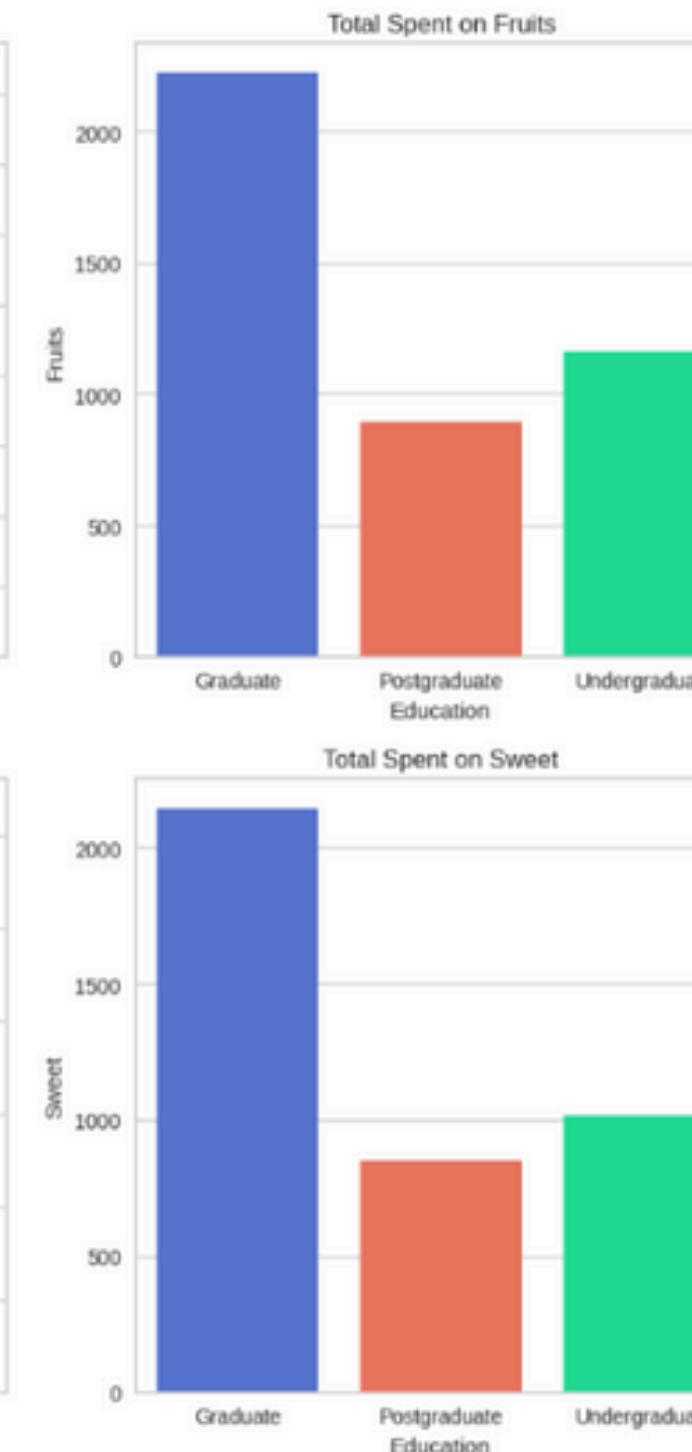
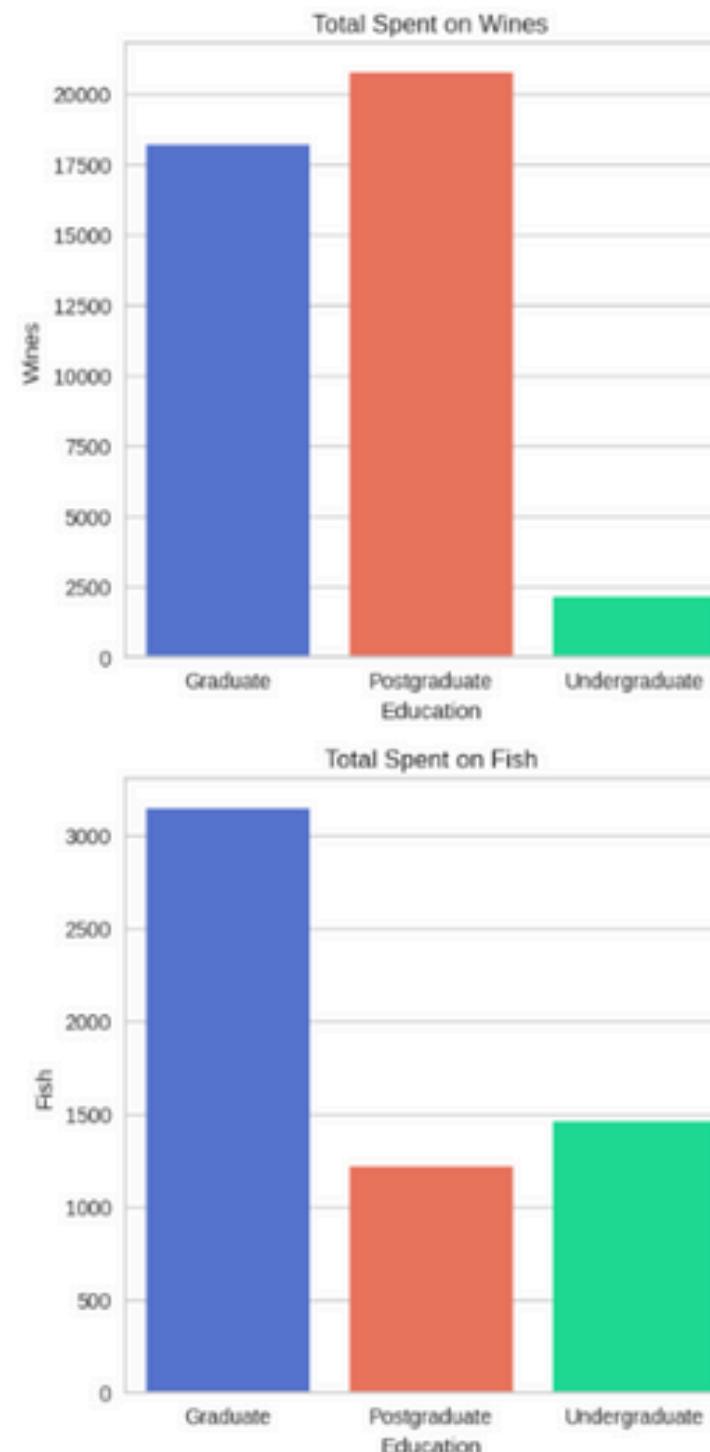
Phần lớn khách hàng thường **mua hàng tại cửa hàng** từ **2 đến 4 lần**. Điều này gợi ý rằng các **cửa hàng vẫn là nơi mua sắm phổ biến hơn** so với web hay catalog.

Số lượt khách hàng truy cập khá **cao** so với số lượt giao dịch qua web, phần lớn khách hàng truy cập từ **5 đến 8 lần** mỗi tháng. Điều này có thể cho thấy **website là điểm tương tác chính** của khách hàng.

Biểu đồ cho thấy có **sự phân bố khá đồng đều về tần suất mua hàng gần đây** của khách hàng. Điều này cho thấy không có sự chênh lệch lớn về mức độ thường xuyên mua sắm giữa các khách hàng. Một số khách hàng mua sắm rất gần đây, trong khi một số khác đã không mua hàng trong một thời gian dài.

PHÂN TÍCH HÀNH VI KHÁCH HÀNG

Tổng chi tiêu cho từng sản phẩm theo từng loại học vấn



Graduate

Sản phẩm nào họ cũng **chi nhiều**, nhiều nhất so với 2 nhóm kia.

Postgraduate

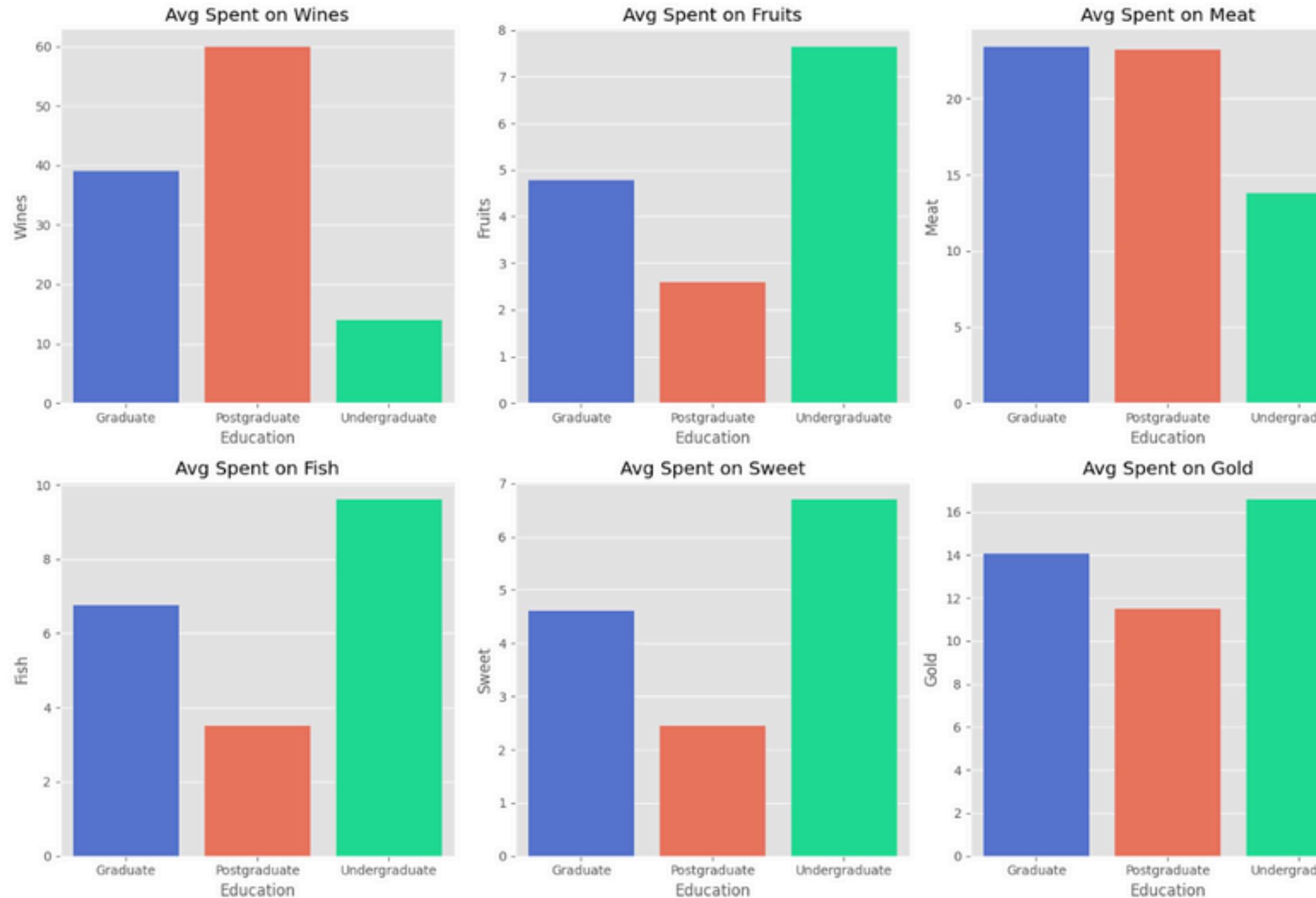
Họ **chủ yếu mua rượu, thịt, sau đó là vàng**. Họ chi khá ít cho cá, hoa quả và đồ ngọt.

Undergraduate

Trừ rượu và thịt ra, tất cả các nhóm khác họ đều **chi tiêu ở mức trung bình** so với 2 nhóm còn lại.

PHÂN TÍCH HÀNH VI KHÁCH HÀNG

Trung bình chi tiêu cho từng sản phẩm theo từng loại **học vấn**



Graduate -----

Trung bình mỗi đơn hàng lượng **tiền chi cho thịt** là **nhiều nhất**.

Postgraduate -----

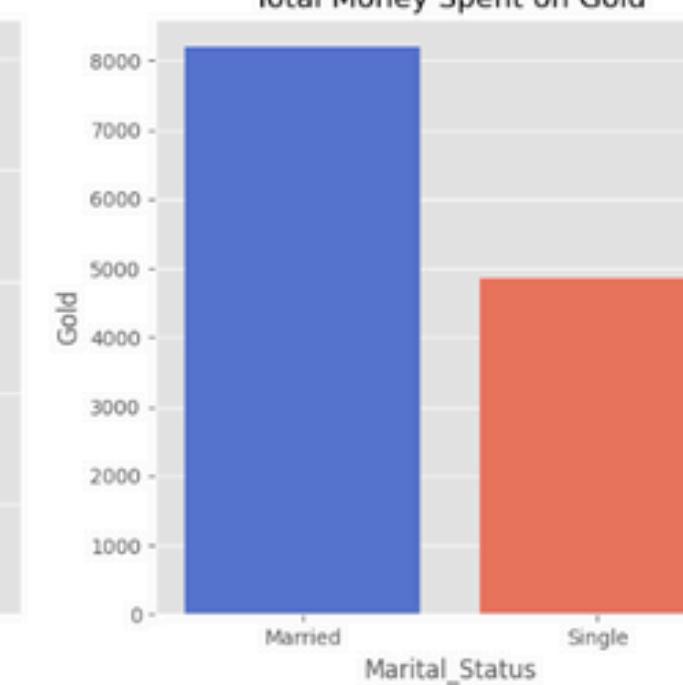
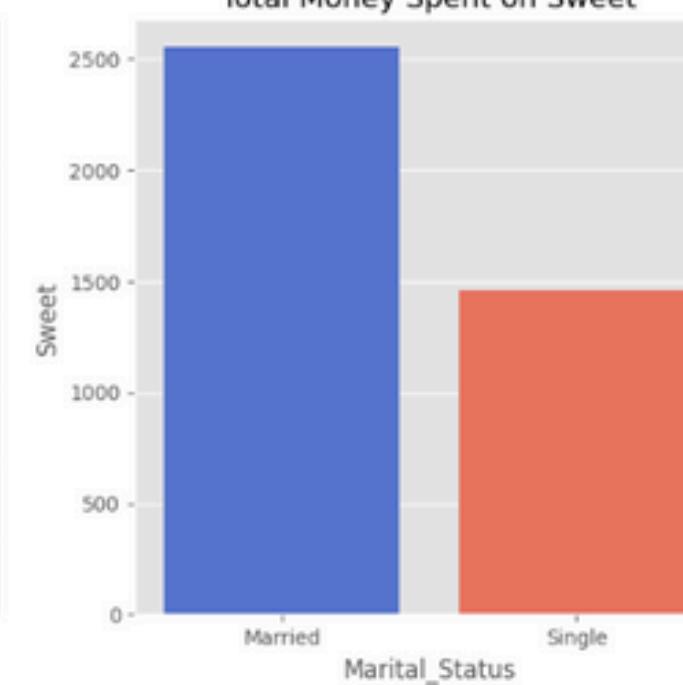
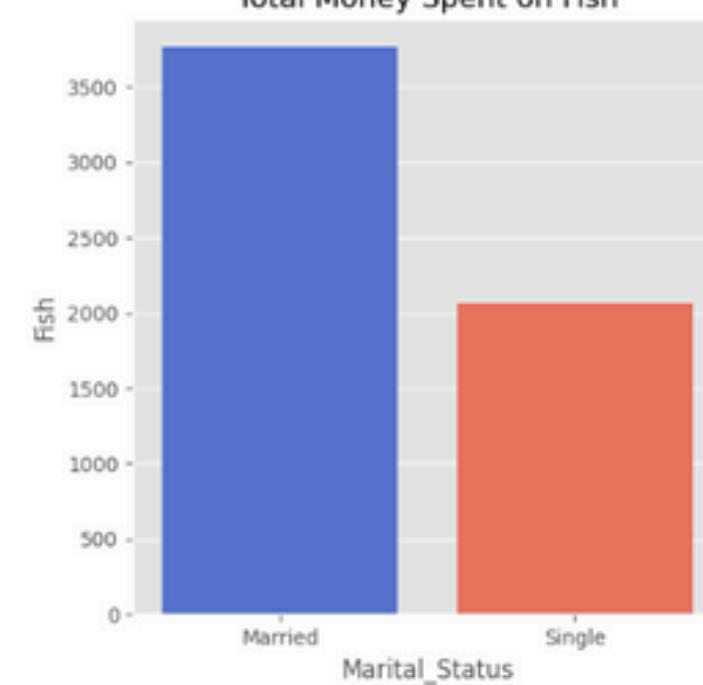
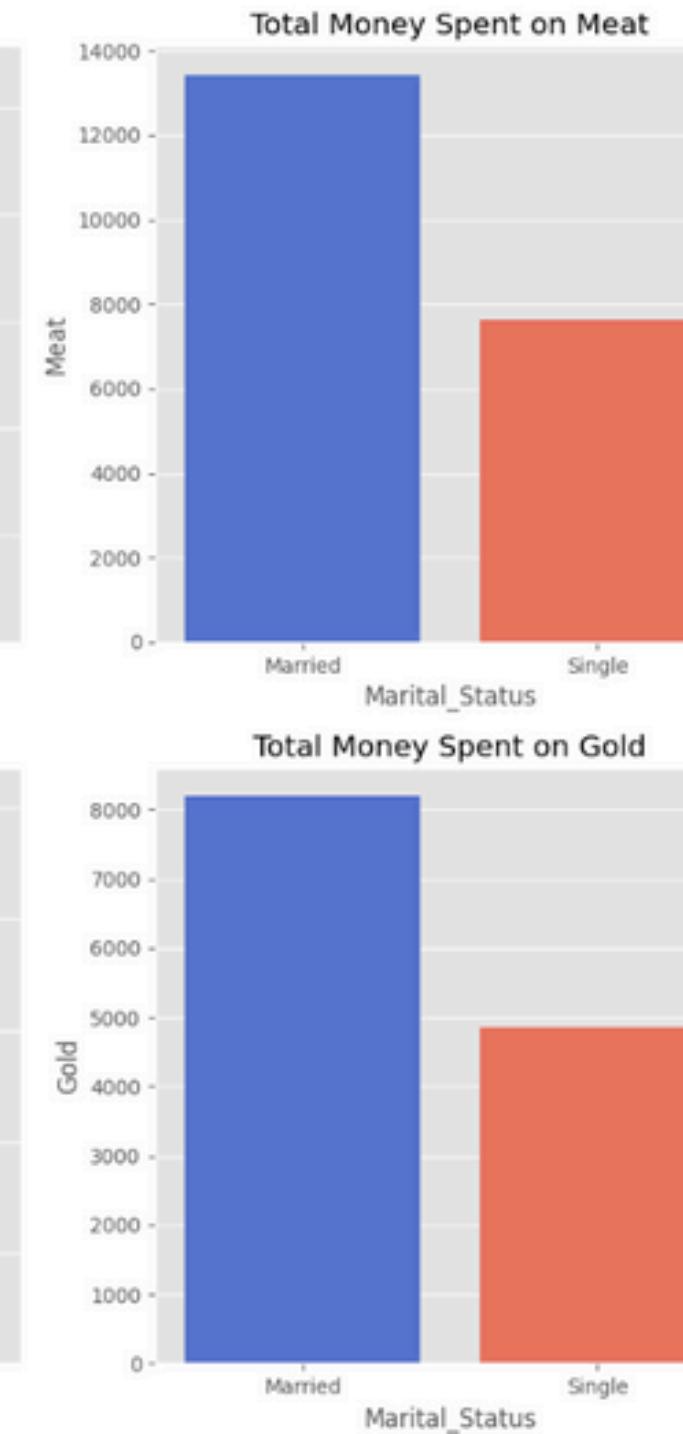
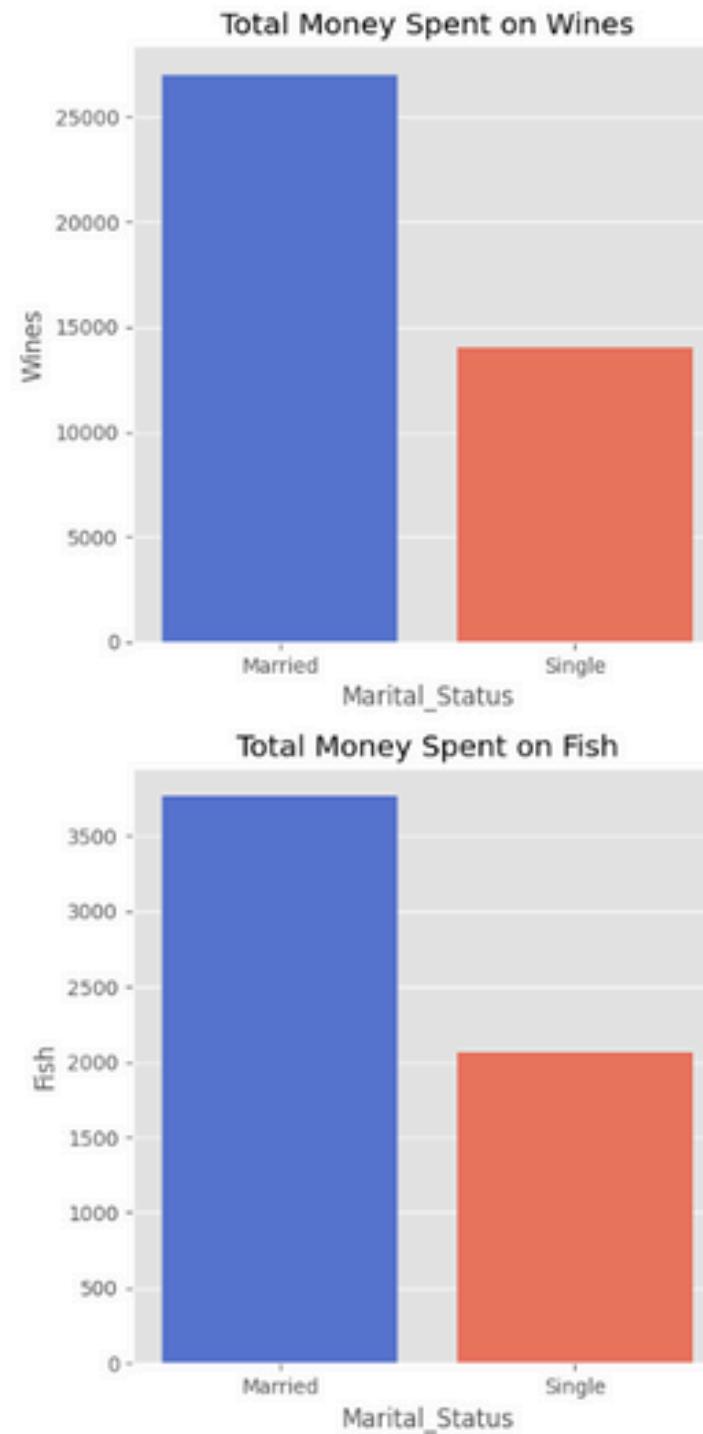
Lượng tiền **chi cho rượu** trung bình mỗi đơn hàng là **nhiều nhất**, sau đó là thịt rồi đến vàng.

Undergraduate -----

Trừ **rượu và thịt** ra, mức tiền **chi tiêu trung bình** cho các nhóm sản phẩm còn lại đều ở mức **cao** so với Graduate và Postgraduate.

PHÂN TÍCH HÀNH VI KHÁCH HÀNG

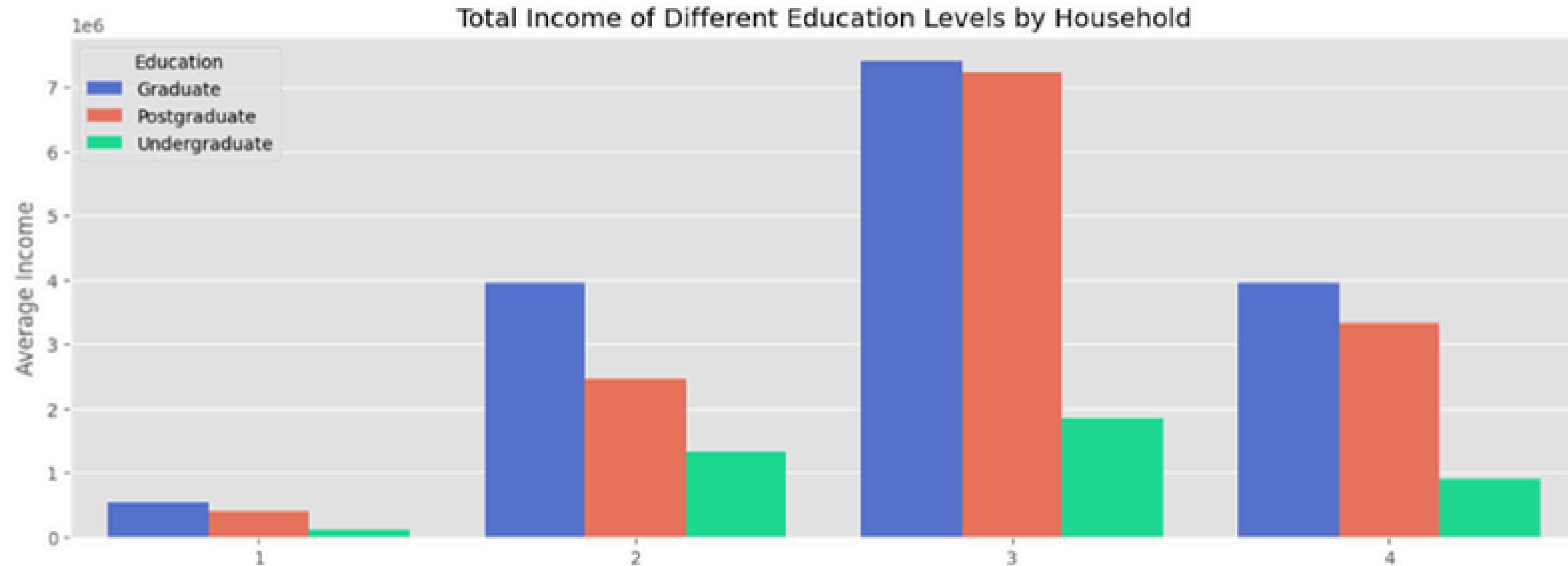
Tổng chi tiêu theo tình trạng hôn nhân



Lượng chi tiêu của nhóm đã kết hôn nhiều gần bằng gấp đôi lượng chi tiêu của những người độc thân.

PHÂN TÍCH HÀNH VI KHÁCH HÀNG

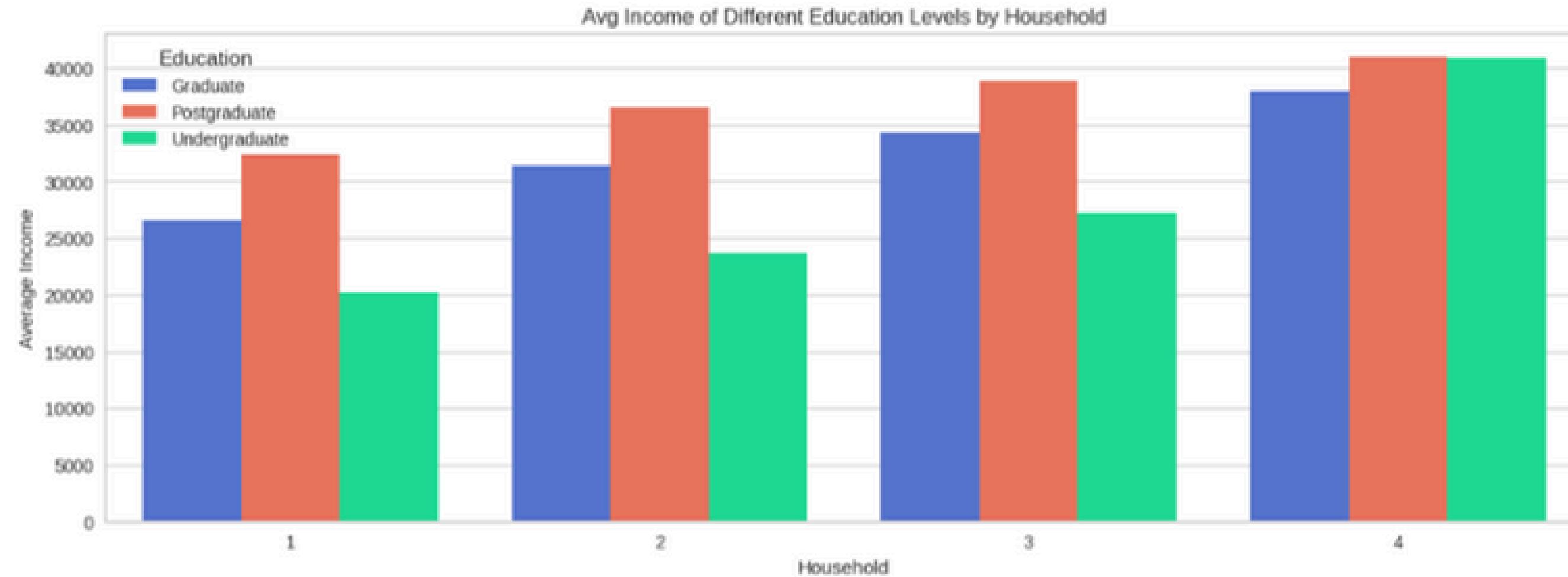
Tổng thu nhập theo số lượng thành viên trong hộ gia đình và trình độ học vấn



Tổng lượng thu nhập của những hộ gia đình có 3 thành viên là cao nhất

PHÂN TÍCH HÀNH VI KHÁCH HÀNG

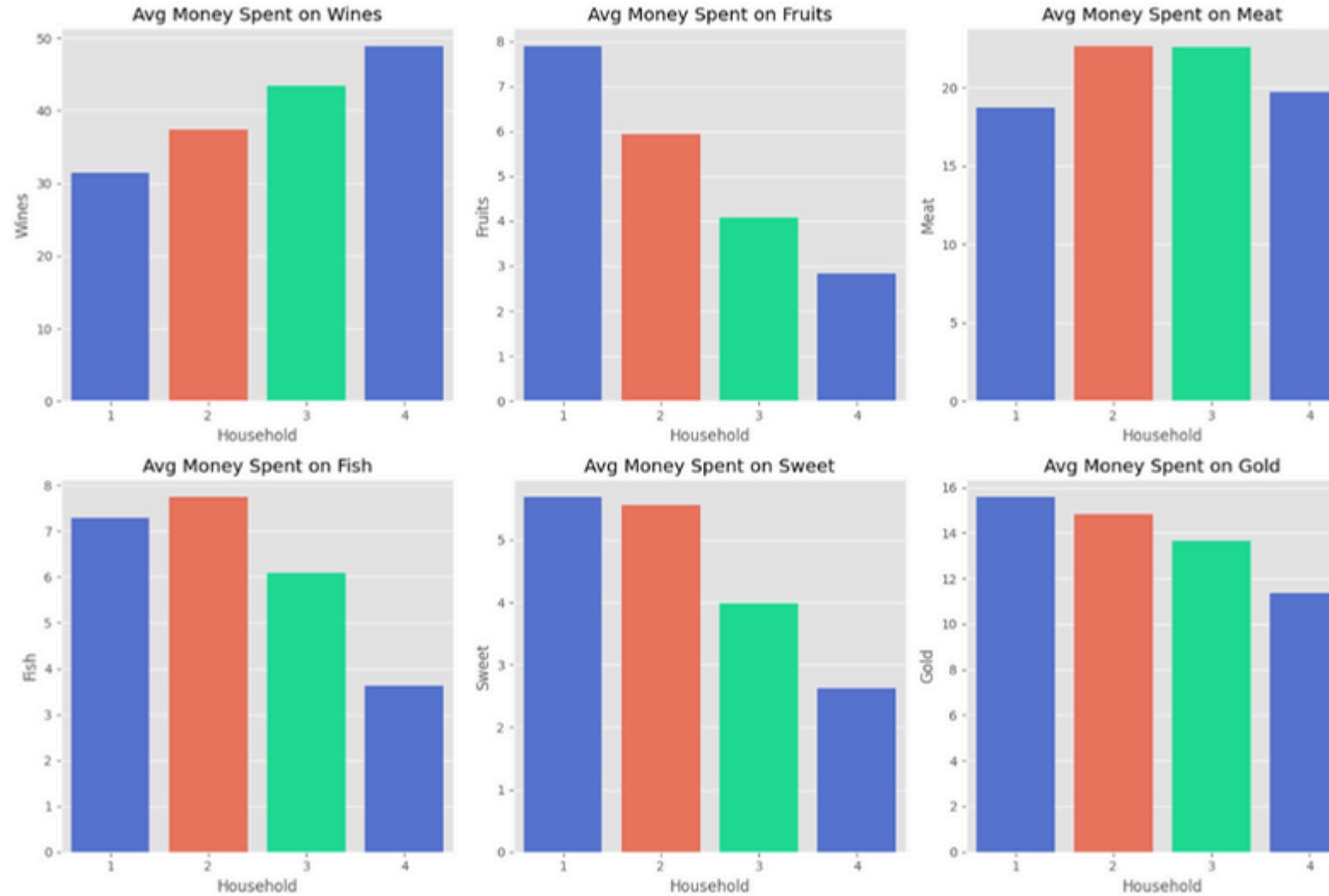
Trung bình thu nhập theo số lượng thành viên trong hộ gia đình và trình độ học vấn



- Đối với hộ **1, 2, 3 thành viên**, trung bình thu nhập của nhóm **Postgraduate** là **cao nhất**, sau đó đến Graduate.
- Với hộ có **1, 2 và 3 thành viên**, mặc dù tổng thu nhập của nhóm Graduate cao hơn Postgraduate nhưng **mức thu nhập trung bình** của **Graduate** lại **thấp hơn**.
- Với hộ **4 thành viên**, trung bình thu nhập của **Postgraduate** gần **bằng** với **Undergraduate**.

PHÂN TÍCH HÀNH VI KHÁCH HÀNG

Trung bình chi tiêu theo số lượng thành viên trong hộ gia đình



- Đối với **rượu**, hộ có **4 thành viên** có lượng chi tiêu trung bình **lớn nhất** so với 3 hộ còn lại.
- Đối với **trái cây** và **vàng** thì lượng chi tiêu trung bình **ngược lại** so với **rượu**.
- Với **thịt** thì lượng chi tiêu trung bình của cả 4 hộ gần như là **ngang nhau**
- Với **cá** và **đồ ngọt** thì hộ **4 người** trung bình chi tiêu **khá ít** so với 3 hộ còn lại

XÂY DỰNG MÔ HÌNH PHÂN CỤM DỰ ĐOÁN

GAUSSIAN MIXTURE MODEL (GMM)

Phân cụm theo **đặc điểm nhân khẩu học**

Thu nhập

Tuổi

Số lượng thành
viên hộ gia đình

K-MEAN

Phân cụm theo RFM

Recency

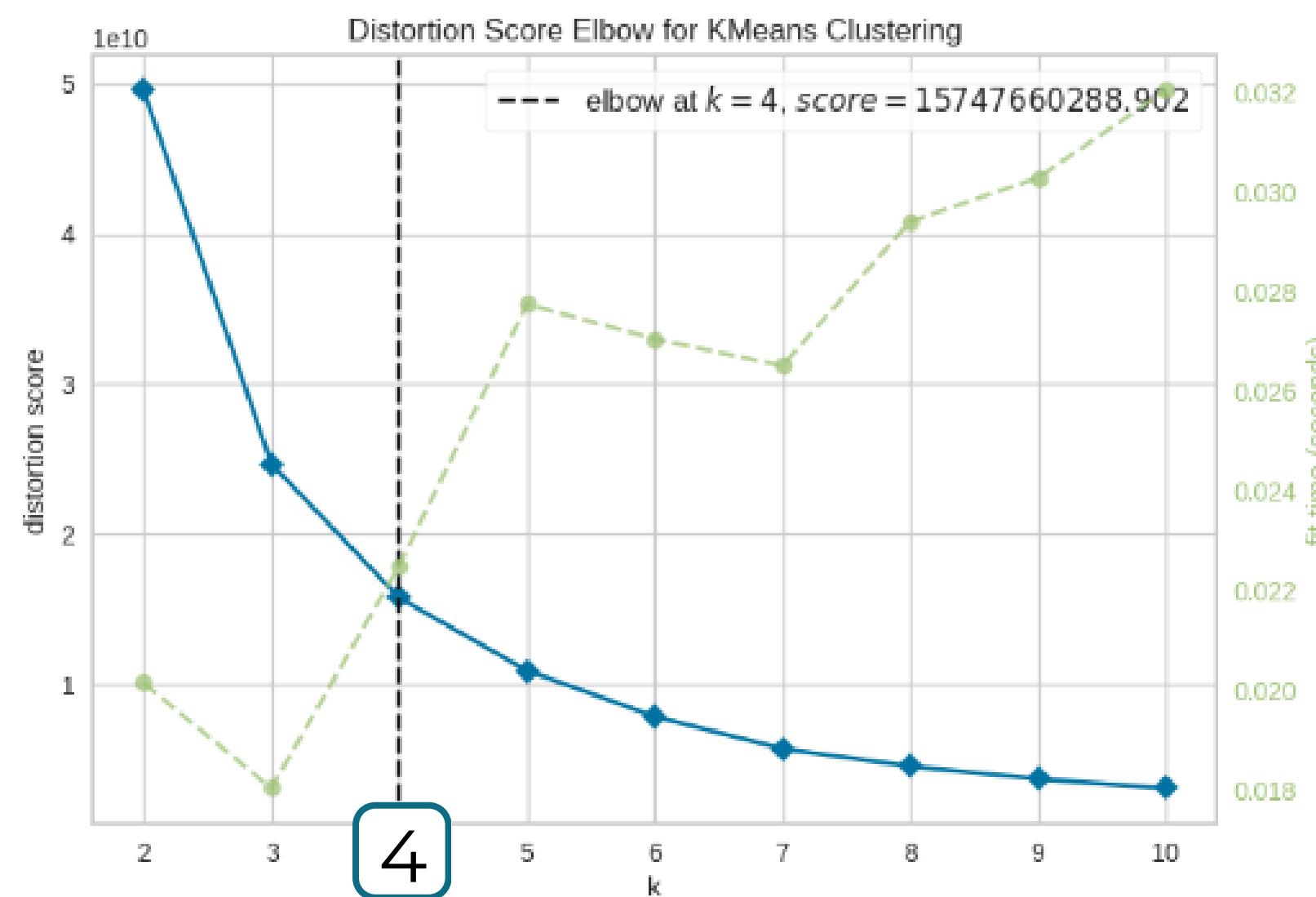
Frequency

Monetary

XÂY DỰNG MÔ HÌNH PHÂN CỤM DỰ ĐOÁN

GAUSSIAN MIXTURE MODEL

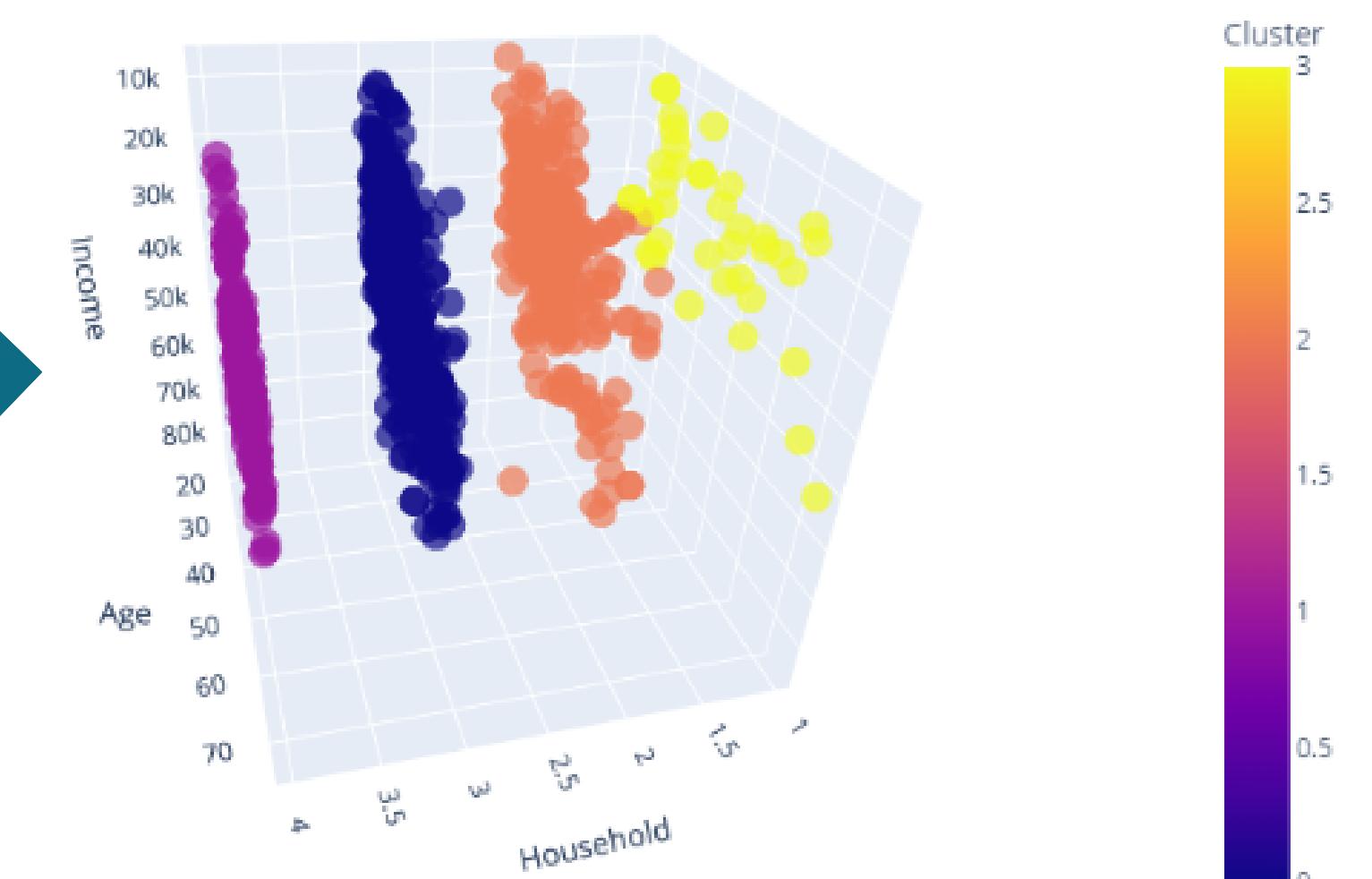
Biểu đồ Elbow



4

Số cụm tối ưu của mô hình là 4 cụm

3D Plot



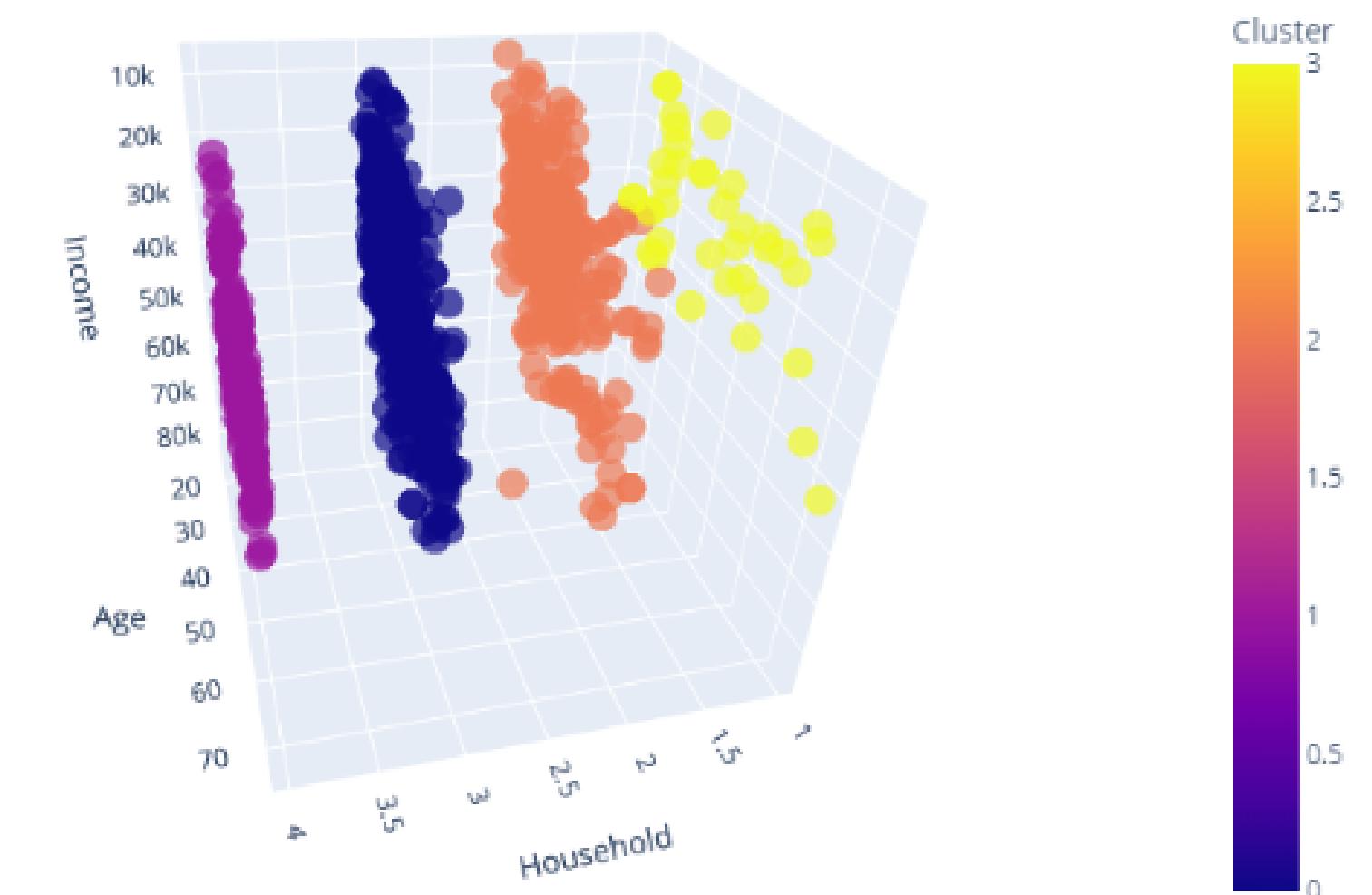
XÂY DỰNG MÔ HÌNH PHÂN CỤM DỰ ĐOÁN

GAUSSIAN MIXTURE MODEL

Số lượng khách hàng theo cụm

Cụm 0	470 khách hàng
Cụm 1	207 khách hàng
Cụm 2	249 khách hàng
Cụm 3	38 khách hàng

3D Plot



ĐẶC ĐIỂM CÁC CỤM

GAUSSIAN MIXTURE MODEL

Đặc điểm nhân khẩu học của các cụm

Cluster	Thu nhập	Tuổi	Số thành viên
Cụm 0	35,212	40	3
Cụm 1	38,452	48	4
Cụm 2	30,093	38	2
Cụm 3	25,323	40	1

Khách hàng có thu nhập thấp (cụm 2 và 3)

Có đặc điểm là có **độ tuổi trẻ hơn** so với 2 cụm còn lại, độ tuổi dưới 40 chiếm số lượng nhiều, **số lượng thành viên trong gia đình từ 1-2 người** trong một hộ.

Khách hàng có thu nhập cao (cụm 0 và 1)

Có đặc điểm là **độ tuổi cao từ 40 - 48** chiếm đa số, **số lượng thành viên từ 3-4 người** trong một hộ.

ĐỀ XUẤT CHIẾN LƯỢC

Đối với khách hàng có thu nhập thấp

- Tập trung vào các **kênh tiếp thị** như **mạng xã hội, email marketing, và tin nhắn SMS**.
- Tạo **nội dung giá trị** với thông điệp liên quan đến **tiết kiệm, giảm giá, và ưu đãi đặc biệt**.
- **Hợp tác** với các **cơ sở thương mại xã hội** để tạo ra các **chương trình khuyến mãi** dành riêng cho cụm khách hàng này.

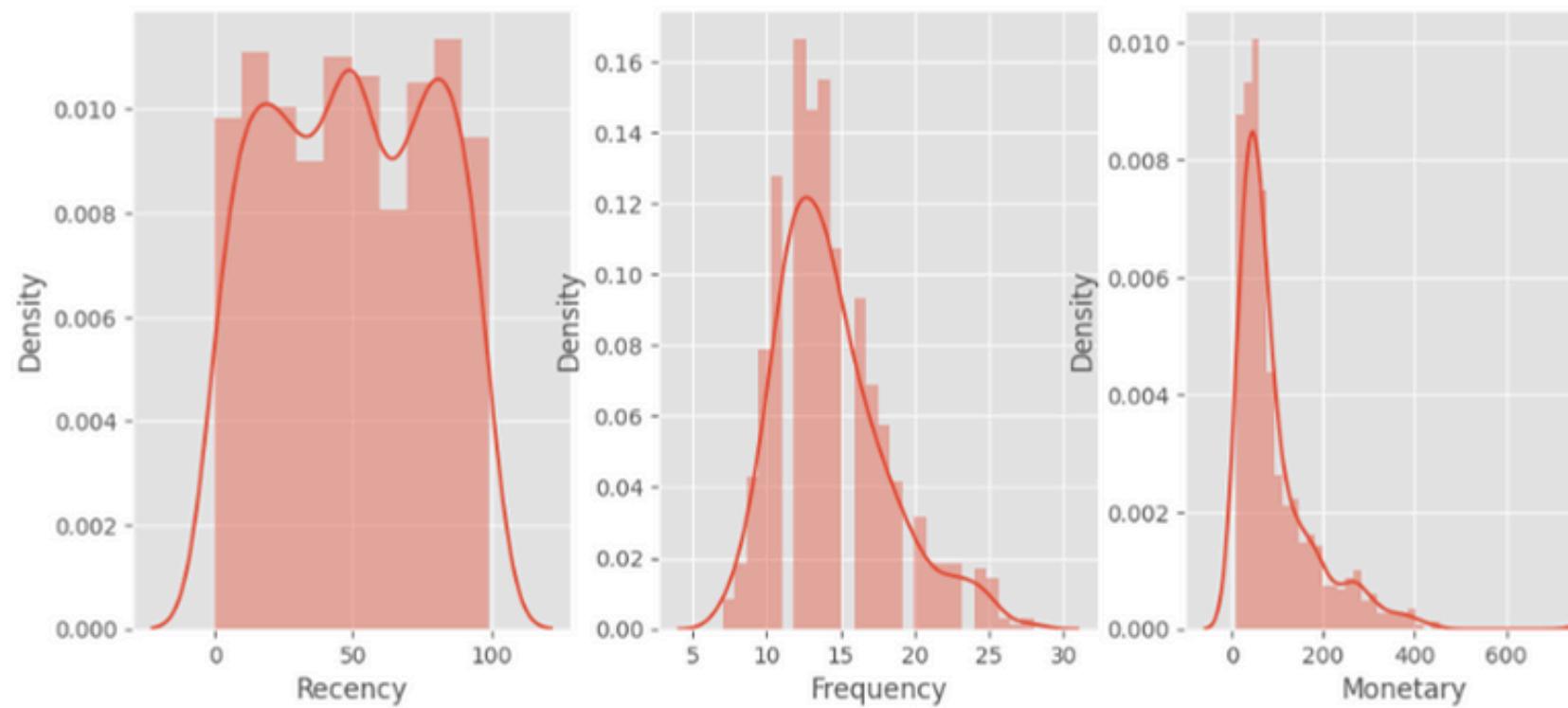
Đối với khách hàng có thu nhập cao

- Sử dụng các **kênh tiếp thị** chuyên nghiệp như **email, sự kiện, và quảng cáo truyền hình**.
- Tạo **nội dung** tập trung vào **chất lượng, thương hiệu, và trải nghiệm sản phẩm**.
- Tạo ra các **chương trình thưởng** dành riêng cho cụm khách hàng này, ví dụ như **chương trình khách hàng thân thiết**.

XÂY DỰNG MÔ HÌNH PHÂN CỤM DỰ ĐOÁN

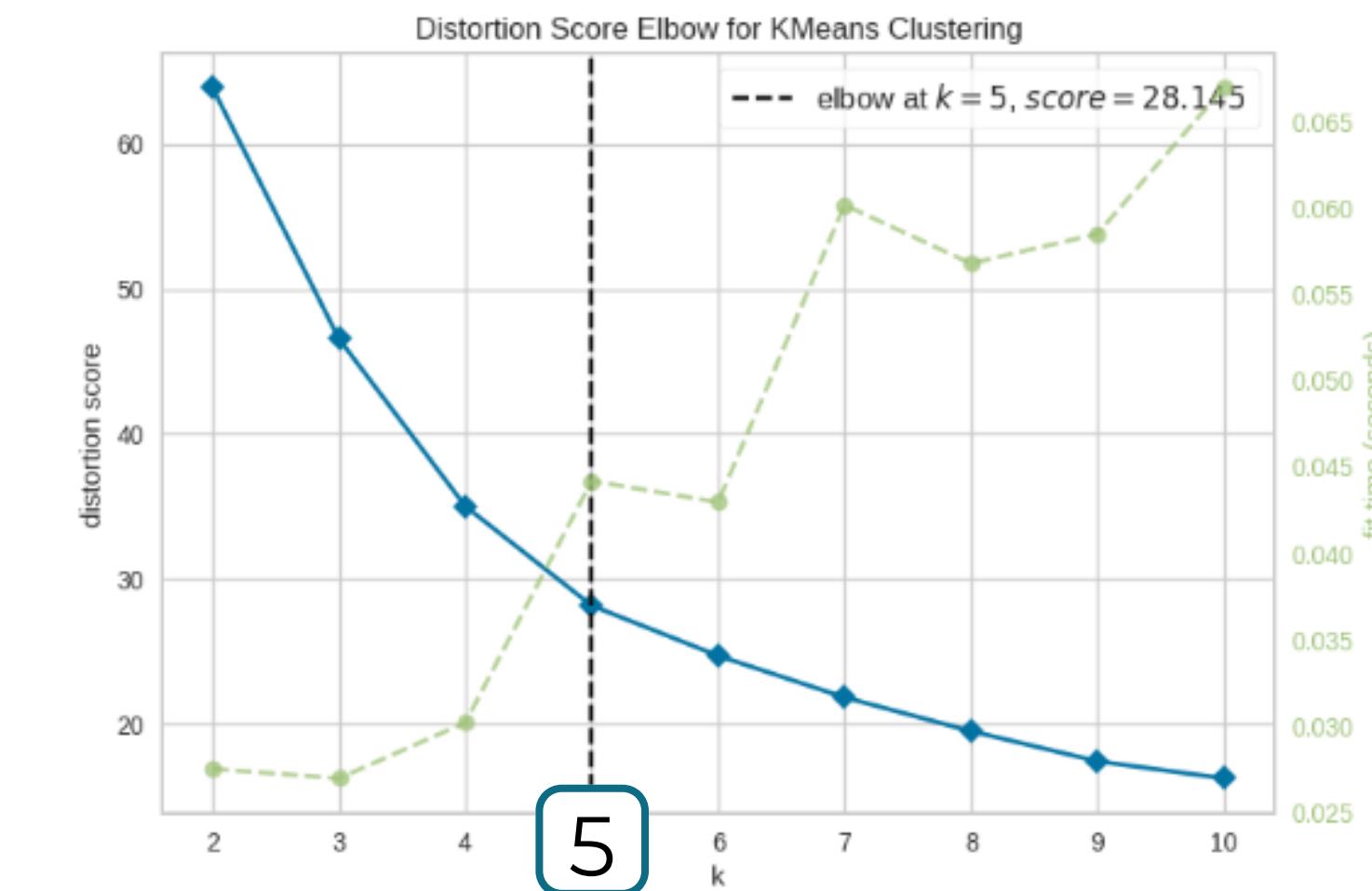
K-MEAN CLUSTERING

Phân phối RFM



- **Thời gian mua hàng gần đây:** Khách hàng có xu hướng mua hàng cách đây **dưới 100 ngày**, và phổ biến nhất là từ **10 - 90 ngày**.
- **Tần suất mua hàng:** Khách hàng có xu hướng mua với tần suất từ **10-15 lần** là nhiều nhất.
- **Giá trị đơn hàng:** Khách hàng có xu hướng chi tiêu với giá trị **tối đa là 200\$**.

Biểu đồ Elbow



Số cụm tối ưu của mô hình là 5 cụm

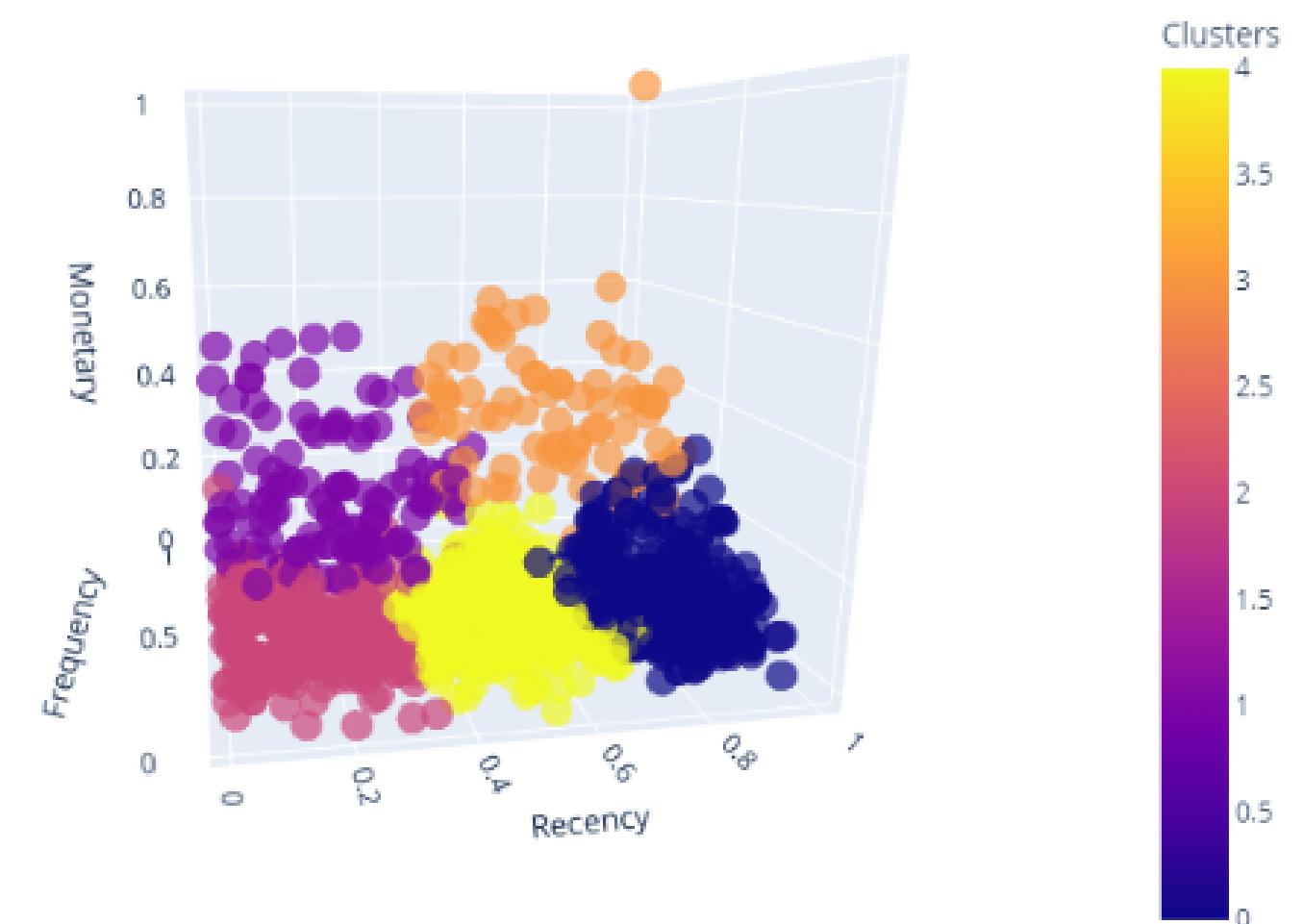
XÂY DỰNG MÔ HÌNH PHÂN CỤM DỰ ĐOÁN

K-MEAN CLUSTERING

Số lượng khách hàng theo cụm

Cụm 0	270 khách hàng
Cụm 1	121 khách hàng
Cụm 2	233 khách hàng
Cụm 3	74 khách hàng
Cụm 4	266 khách hàng

3D Plot



ĐẶC ĐIỂM CÁC CỤM

K-MEAN CLUSTERING

Cluster	Recency	Frequency	Monetary
Cụm 0	0.84	0.31	0.09
Cụm 1	0.23	0.56	0.23
Cụm 2	0.15	0.26	0.06
Cụm 3	0.73	0.69	0.36
Cụm 4	0.50	0.28	0.06

Cụm 0

Recency cao (0.84), **Tần suất mua hàng của họ thấp** (0.31), đồng thời **chi tiêu ít** (0.09).

→ **Mục tiêu của các chiến dịch khuyến mãi mới.**

Cụm 1

Recency thấp (0.23), **tần suất mua hàng cao** (0.56) và **số tiền chi tiêu trung bình** (0.23).

→ **Khách hàng trung thành**.

Cụm 2

Cụm này có **Recency thấp** (0.15), **tần suất mua hàng thấp** (0.26) và **số tiền chi tiêu ít** (0.06).

→ **Khách hàng tiềm năng**.

Cụm 3

Cụm này có **Recency cao** (0.73), **tần suất mua hàng cao** (0.69) và **số tiền chi tiêu khá lớn** (0.36).

→ **Khách hàng quan trọng**.

Cụm 4

Cụm này có **Recency trung bình** (0.50), **tần suất mua hàng thấp** (0.28) và **số tiền chi tiêu ít** (0.06).

ĐỀ XUẤT CHIẾN LƯỢC

CỤM 0

Tạo các **chiến dịch khuyến mãi mới** để thu hút khách hàng quay lại mua sắm.

Giảm giá hoặc **tặng quà**.

Tạo **chương trình thưởng** dành riêng cho họ.

CỤM 1

Duy trì mối quan hệ với họ bằng cách **gửi email thông tin sản phẩm mới, ưu đãi**.

Tạo **chương trình khách hàng trung thành**.

CỤM 2

Khuyến mãi cho lần mua hàng **đầu tiên**.

Tạo **nội dung giới thiệu sản phẩm** trên **mạng xã hội**.

CỤM 3

Duy trì mối quan hệ thân thiết với họ thông qua **sự kiện, chương trình khách hàng VIP**.

Tạo **nội dung chất lượng** về sản phẩm và thương hiệu để **duy trì sự quan tâm** của họ.

CỤM 4

Tạo **chương trình khuyến mãi đặc biệt** cho lần **mua hàng tiếp theo**.

Gửi thông báo về **sản phẩm mới** và **ưu đãi** qua **email**.



THANK YOU
FOR YOUR ATTENTION