

**BỘ GIÁO DỤC VÀ ĐÀO TẠO**

**TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT TP.HCM**

sdfadf

**Thành Phố Hồ Chí Minh, tháng 5 năm 2023**

**BÁO CÁO CUỐI KỲ MÔN TƯƠNG TÁC DỮ LIỆU TRỰC QUAN**

**ĐỀ TÀI**

**MARKETING ANALYTICS**

******

**GVHD: ThS Lê Minh Tân**

**Lớp: Thứ 4 ( tiết 1-4 )**

**Sinh viên thực hiện: Nhóm 4**

**Phan Quốc Lưu MSSV: 20133065**

**Cao Trọng Nghĩa MSSV:20133071**

**DANH SÁCH NHÓM LÀM ĐỒ ÁN CUỐI KỲ**

**MÔN HỌC: TƯƠNG TÁC DỮ LIỆU TRỰC QUAN**

**HỌC KỲ II – NĂM HỌC: 2022-2023**

1. **Mã môn học: IDVI333677\_22\_2\_02**
2. **Giảng viên hướng dẫn: Ths. Lê Minh Tân**
3. **Tên đề tài: Marketing Analytics**
4. **Danh sách nhóm làm đồ án cuối kỳ**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **STT** | **Họ và tên sinh viên** | **MSSV** | **% Hoàn Thành** | **Kí Tên** |
| 1 | Phan Quốc Lưu | 20133065 | 100% |  |
| 2 | Cao Trọng Nghĩa | 20133071 | 100% |  |

* **Tỉ lệ hoàn thành: 100%**
* **Trưởng nhóm: Phan Quốc Lưu*Nhận xét của giảng viên:***……………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………….………………………………………………………

*Tháng 05 năm 2023*  
**Giảng viên chấm điểm**

**LỜI CẢM ƠN**

Lời đầu tiên, chúng em xin gửi lời cảm ơn chân thành đến Thầy đã hướng dẫn, giúp đỡ và chỉ dẫn chúng em trong quá trình thực hiện đề tài này. Thầy đã tận tình chia sẻ kiến thức và kinh nghiệm của mình, giúp chúng em hiểu sâu hơn về đề tài. Thầy cũng đã giúp đỡ chúng em giải quyết những khó khăn và trở ngại trong quá trình nghiên cứu, giúp chúng em hoàn thành đề tài này. Không có sự hỗ trợ đó, chúng em không thể hoàn thành đề tài một cách thành công.

Chúng em cũng xin gửi lời cảm ơn đến các thành viên trong nhóm. Mỗi thành viên đã đóng góp một phần công sức, tài năng và thời gian để hoàn thành đề tài này. Chúng em đã học hỏi và trau dồi thêm nhiều kỹ năng, kiến thức mới từ nhau. Đặc biệt, chúng em cảm ơn vì mối quan hệ tình cảm và sự đoàn kết của nhóm trong suốt quá trình làm việc.

Một lần nữa, chúng em xin chân thành cảm ơn tất cả những người đã giúp đỡ chúng em trong quá trình nghiên cứu và thực hiện đề tài này. Sự hỗ trợ của các bạn đã làm cho đề tài của chúng em trở nên hoàn hảo hơn

**LỜI MỞ ĐẦU**

Đề tài "Phân tích dữ liệu thị trường" là một đề tài rất thú vị và có tính ứng dụng cao trong thực tế. Thị trường luôn thay đổi và có nhiều yếu tố ảnh hưởng đến giá cả và các hoạt động kinh doanh. Do đó, việc phân tích và đánh giá dữ liệu thị trường là rất quan trọng để các doanh nghiệp có thể đưa ra các quyết định thích hợp và hiệu quả.

Trong đề tài này, chúng ta sẽ sử dụng dữ liệu thị trường được cung cấp bởi Kaggle để thực hiện phân tích và đánh giá. Dữ liệu này bao gồm thông tin về khách hàng, sản phẩm và chiến dịch tiếp thị của một doanh nghiệp. Chúng ta sẽ áp dụng các kỹ thuật phân tích dữ liệu như trực quan hóa dữ liệu, phân tích hồi quy và phân tích nhân tố để tìm hiểu và giải thích mối quan hệ giữa các biến trong dữ liệu.

Mục đích của đề tài này là tìm hiểu cách thức hoạt động của thị trường và các yếu tố ảnh hưởng đến hoạt động kinh doanh. Kết quả của đề tài này sẽ giúp các doanh nghiệp có thể đưa ra các quyết định thông minh và hiệu quả trong hoạt động của mình. Chúng em cũng hy vọng rằng đề tài này sẽ giúp các bạn hiểu thêm về phân tích dữ liệu và ứng dụng của nó trong thực tiễn.

**Mục Lục**

[**I.** **Giới thiệu dữ liệu** 6](#_Toc134261925)

[**II.** **Xử lí dữ liệu** 8](#_Toc134261926)

[**III.** **Các đồ thị / biểu đồ** 9](#_Toc134261927)

[**1.** **Tỷ lệ khách hàng có trẻ em đã mua hàng theo năm sinh** 9](#_Toc134261928)

[**2.** **Mức chi tiêu cá nhân dựa vào tổng số lần mua sản phẩm theo trình độ học vấn**………………………………………………………………………………...9](#_Toc134261929)

[**3.** **Tỉ lệ phân loại học vấn của khách hàng** 10](#_Toc134261930)

[**4.** **Mức độ tiêu thụ sản phẩm theo danh mục và theo từng vùng** 11](#_Toc134261931)

[**5.** **Mối quan hệ giữa thu nhập và chi phí cho các sản phẩm** 12](#_Toc134261932)

[**6.** **Biểu đồ so sánh tỉ lệ chấp nhận chiến dịch tiếp thị theo vùng** 14](#_Toc134261933)

[**7.** **Biểu đồ nâng cao** 15](#_Toc134261934)

[**7.1. Thu nhập của khách hàng dựa trên trình độ học vấn** 15](#_Toc134261935)

[**7.2. Lưới chọn thuộc tính** 16](#_Toc134261936)

[**8.** **Đường trung bình thu nhập theo năm sinh** 18](#_Toc134261937)

[**9.** **Mối quan hệ giữa thu nhập trung bình và số lần mua rượu vang theo vùng** 19](#_Toc134261938)

[**10.** **Các công nghệ/phần mềm sử dụng** 19](#_Toc134261939)

[**IV.** **Kết luận** 20](#_Toc134261940)

[**TÀI LIỆU THAM KHẢO** 21](#_Toc134261941)

**ĐỀ TÀI : MARKETING ANALYST**

1. **Giới thiệu dữ liệu**

**Marketing\_data**

Bộ dữ liệu "market-data-analysis" trên Kaggle là một tập hợp các thông tin về chiến dịch tiếp thị của một công ty trong năm 2012. Tập dữ liệu này chứa thông tin về khách hàng, chiến dịch tiếp thị, các thông số thị trường và doanh số bán hàng.

Tập dữ liệu này bao gồm một file CSV với tên "marketing\_data.csv". File này có 2,240 dòng và 28 cột, bao gồm các thông tin về tuổi, giới tính, hôn nhân, thu nhập, giáo dục, số lần liên hệ, loại khách hàng, thông tin về chiến dịch tiếp thị và các thông số thị trường.

Nguồn:<https://www.kaggle.com/code/hamburger22/market-data-analysis/input?select=marketing_data.csv>

Mô tả:

* + - * ID: Mã khách hàng
      * Year\_Birth: Năm sinh của khách hàng
      * Education: Trình độ học vấn của khách hàng
      * Marital\_Status: Tình trạng hôn nhân
      * Income: Thu nhập hằng năm
      * Kidhome: Số trẻ em trong gia đình
      * Teenhome: Số thanh thiếu niên trong gia đình
      * Dt\_Customer: Ngày đăng ký của khách hàng với công ty
      * Recency: Số ngày kể từ lần mua hàng cuối cùng
      * MntWines: Chi phí cho rượu vang
      * MntFruits: Chi phí cho trái cây
      * MntMeatProducts: Chi phí cho thịt
      * MntFishProducts: Chi phí cho hải sản
      * MntSweetProducts: Chi phí cho đồ ngọt
      * MntGoldProds: Chi phí cho vàng
      * NumDealsPurchases: Số lần mua hàng được giảm giá
      * NumWebPurchases: Số lần mua hàng trên web của cửa hàng
      * NumCatalogPurchases: Số lần mua hàng được thực hiện qua bảng danh mục những mặt hàng
      * NumStorePurchases: Số lần mua hàng tại cửa hàng
      * NumWebVistisMonth: Số lượt truy cập vào trang web của công ty trong tháng trước
      * AcceptedCmp1: Chấp nhận đề nghị chiến dịch thứ 1 (1 nếu khách hàng chấp nhận, 0 là ngược lại)
      * AcceptedCmp2: Chấp nhận đề nghị chiến dịch thứ 2 (1 nếu khách hàng chấp nhận, 0 là ngược lại)
      * AcceptedCmp3: Chấp nhận đề nghị chiến dịch thứ 3 (1 nếu khách hàng chấp nhận, 0 là ngược lại)
      * AcceptedCmp4: Chấp nhận đề nghị chiến dịch thứ 4 (1 nếu khách hàng chấp nhận, 0 là ngược lại)
      * AcceptedCmp5: Chấp nhận đề nghị chiến dịch thứ 5 (1 nếu khách hàng chấp nhận, 0 là ngược lại)
      * Response: Chấp nhận đề nghị chiến dịch cuối cùng (1 nếu khách hàng chấp nhận, 0 là ngược lại)
      * Complain: Lời phần nàn trong 2 năm qua
      * Country: Vùng, khu vực

1. **Xử lí dữ liệu**

**Xử lí dữ liệu trong file marketing\_data.csv**

* + - Xử lí khoảng trắng trong cột
    - Loại bỏ các kí tự đặc biệt trong cột Income
    - Xóa các giá trị ngoại lệ trong cột Income
    - Định dạng lại giá trị của cột Dt\_Customer thành kiểu datetime
    - Thay thế giá trị trống của cột Income bằng giá trị trung vị của cột đó
    - Loại bỏ 2 cột ID và Marital\_Status
    - Tính tổng chi tiêu của các cột liên quan đến chi tiêu và thêm vào cột 'MntSpent'
    - Tính tổng số lần mua của khách hàng từ các cột liên quan đến mua hàng và thêm vào cột 'NumPurchases'
    - Tính tổng số trẻ em của khách hàng từ các cột liên quan đến số lượng trẻ em và thêm vào cột 'TotalKid'
    - Tính tổng số chương trình khuyến mại đã chấp nhận của khách hàng từ các cột liên quan đến chương trình khuyến mại và thêm vào cột 'TotalAccepted'.
    - Thay đổi Year\_Birth thành tuổi Age



1. **Các đồ thị / biểu đồ**
2. **Tỷ lệ khách hàng có trẻ em đã mua hàng**

**Chart, pie chart

Description automatically generated**

**Chart, pie chart

Description automatically generated**

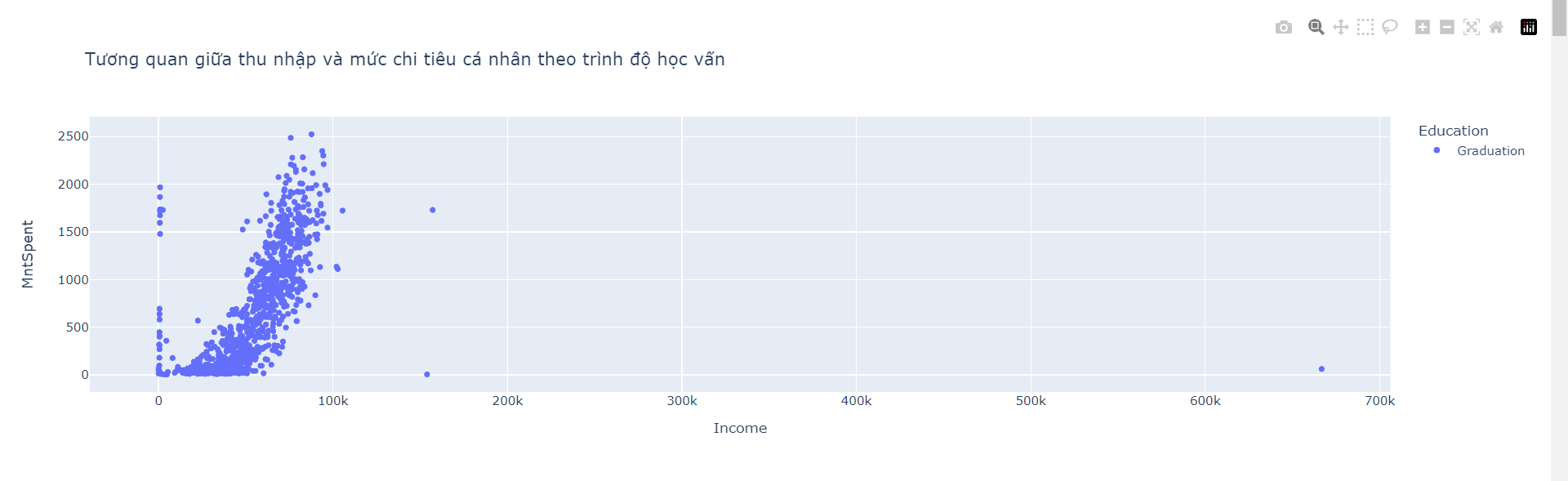
**Chart, pie chart

Description automatically generated**

Biểu đồ trên thể hiện tỷ lệ khách hàng đã mua sản phẩm (product\_spent\_percent) và tỷ lệ khách hàng có trẻ em hay không (kidhome\_percent) dựa trên dữ liệu năm sinh mà mình nhập vào

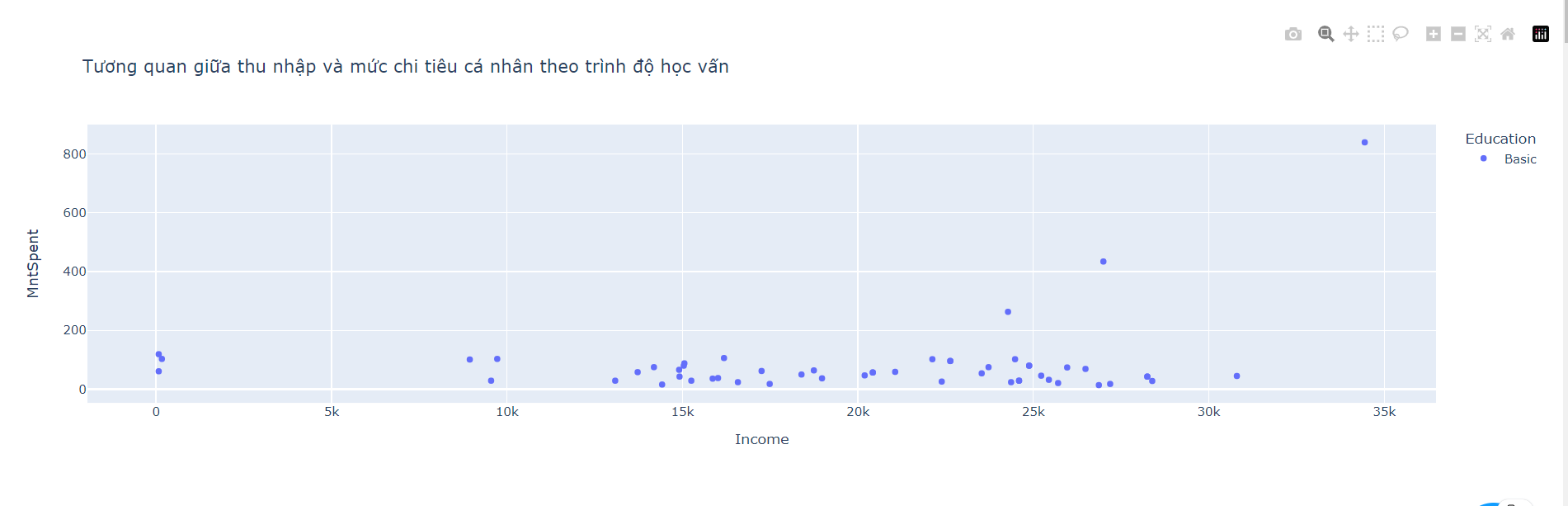
Vì vậy, ta có thể quan sát rằng trong tập dữ liệu khách hàng được phân tích, không quá nhiều khách hàng có trẻ em (chỉ 39.6%), nhưng tổng số tiền mà khách hàng đã chi tiêu cho các loại sản phẩm lại khá cao (chiếm 60.4%). Điều này cho thấy nhóm khách hàng này là những người tiêu dùng tích cực và sẵn sàng chi tiêu nhiều tiền cho các sản phẩm.

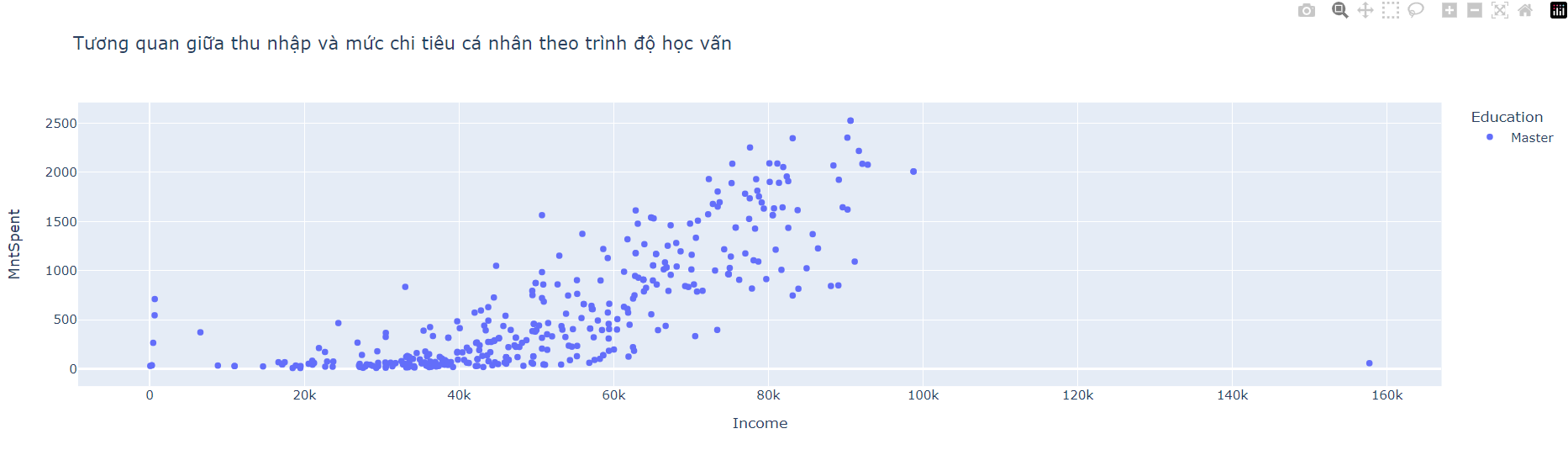
1. **Mức chi tiêu cá nhân dựa vào tổng số lần mua sản phẩm theo trình độ học vấn**

****

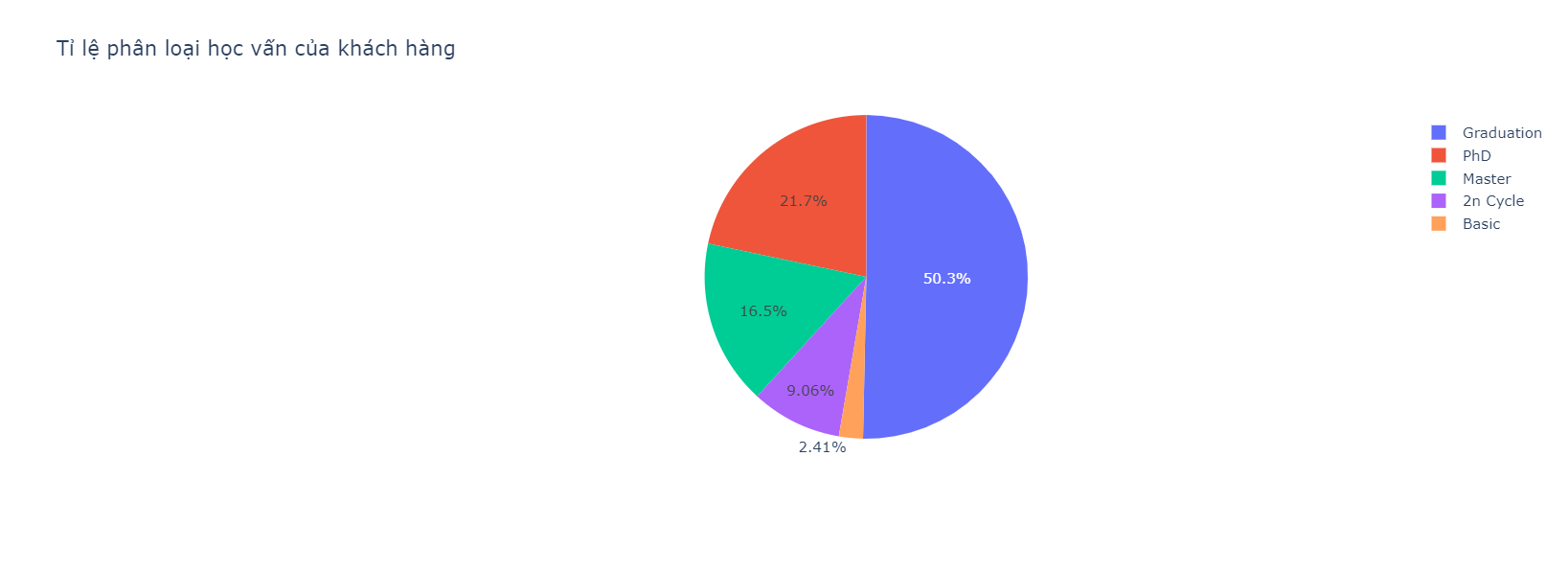
Biều đồ scatter thể hiện độ tương quan giữa thu nhập (Income) và tổng số tiền đã chi tiêu cho sản phẩm (MntSpent) của khách hàng, được phân loại theo trình độ học vấn (Education). Biểu đồ cho thấy sự phân bố của các điểm dữ liệu trên mặt phẳng 2 chiều, trong đó mỗi điểm biểu thị một khách hàng, với trục hoành là thu nhập và trục tung là tổng số tiền đã chi tiêu.

Có thể xem mức độ thu nhập và tổng số tiền đã chi tiêu cho sản phẩm theo từng loại học vấn bằng cách lựa chọn trình độ học vấn để xem





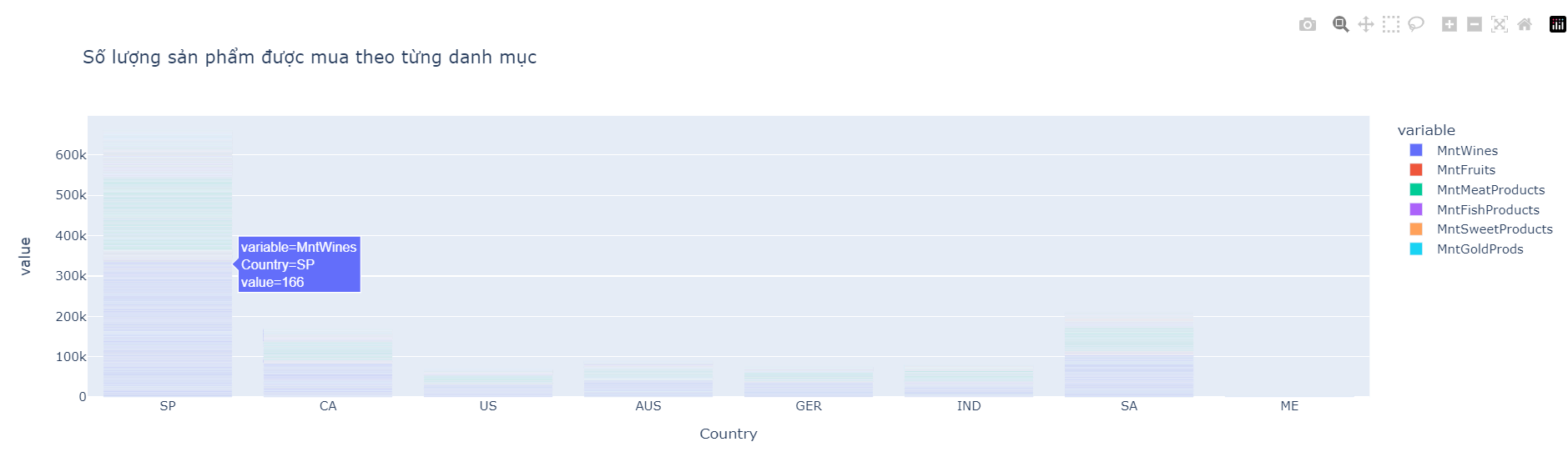
1. **Tỉ lệ phân loại học vấn của khách hàng**

****

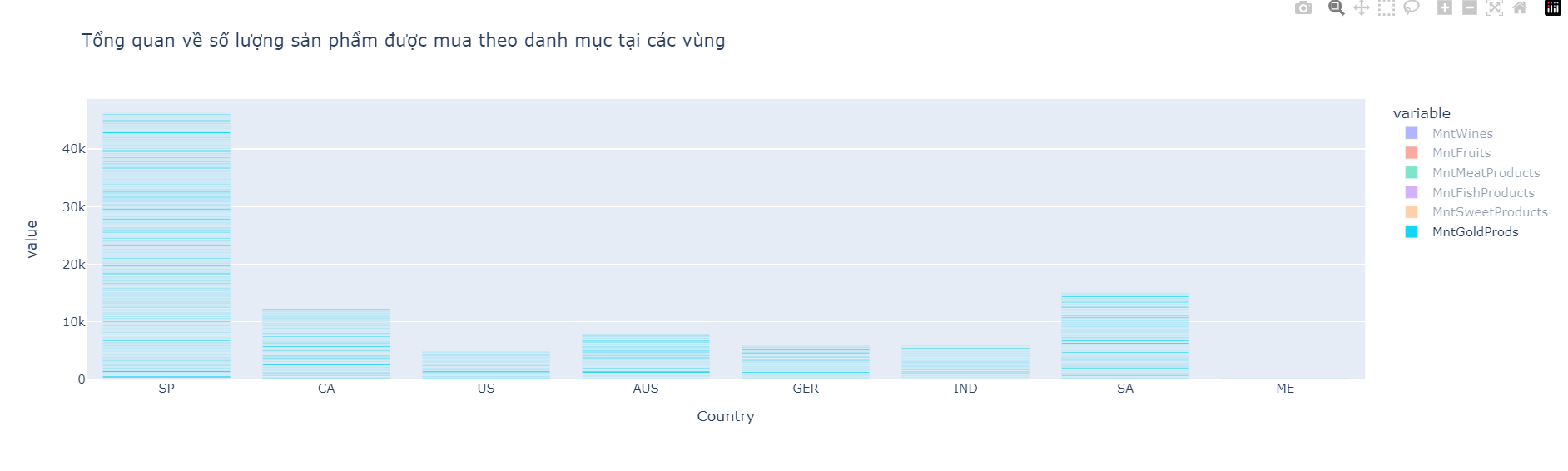
Dựa trên biểu đồ ta có thể phân loại trình độ học vấn của khách hàng

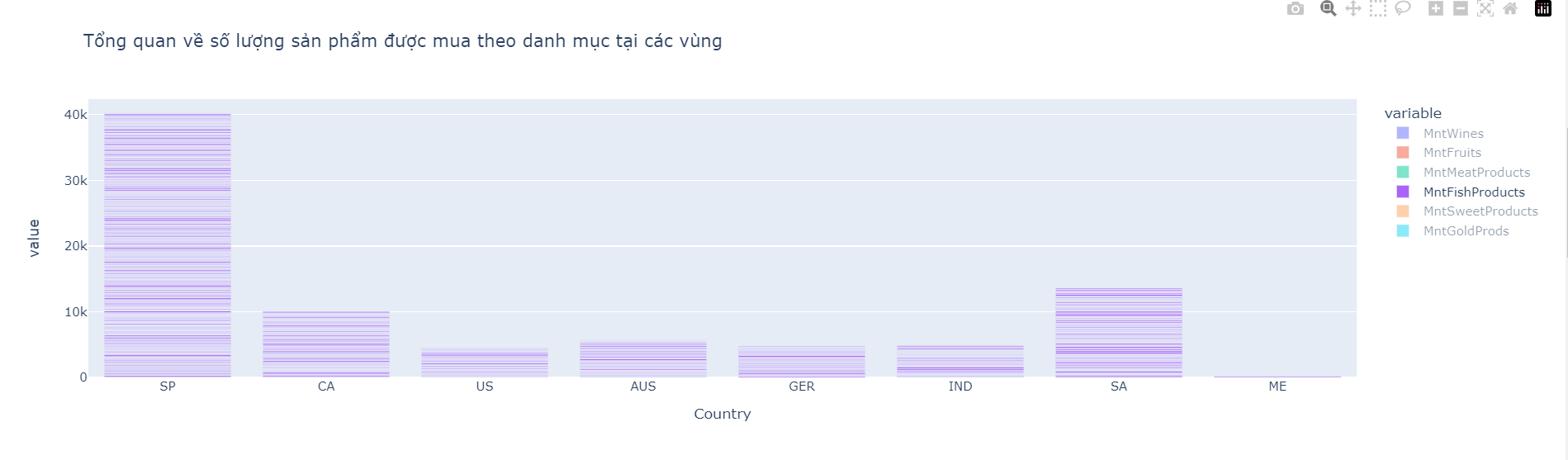
* Graduation: 50.3%
* PhD: 21.7%
* Master: 16.5%
* 2n Cycle: 9.06%
* Basic: 2.41%

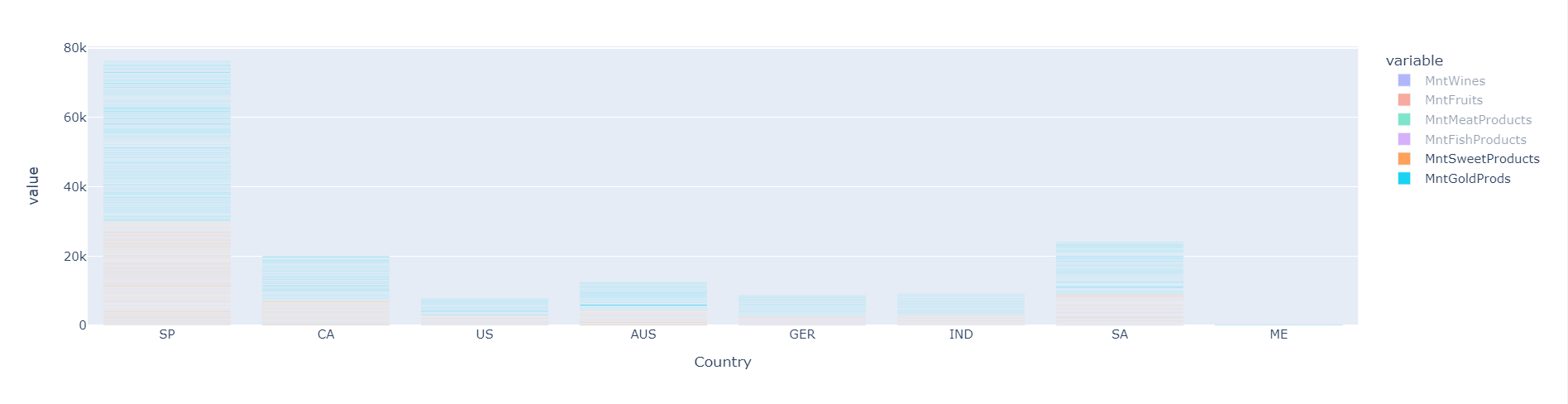
1. **Mức độ tiêu thụ sản phẩm theo danh mục và theo từng vùng**

****

Biểu đồ trên là một biểu đồ cột (bar chart) thể hiện số lượng sản phẩm được mua theo từng danh mục, với trục x là tên các quốc gia, trục y là số lượng sản phẩm được mua. Các danh mục sản phẩm được mua bao gồm: rượu vang (MntWines), trái cây (MntFruits), sản phẩm thịt (MntMeatProducts), sản phẩm hải sản (MntFishProducts), sản phẩm đồ ngọt (MntSweetProducts) và sản phẩm vàng (MntGoldProds)

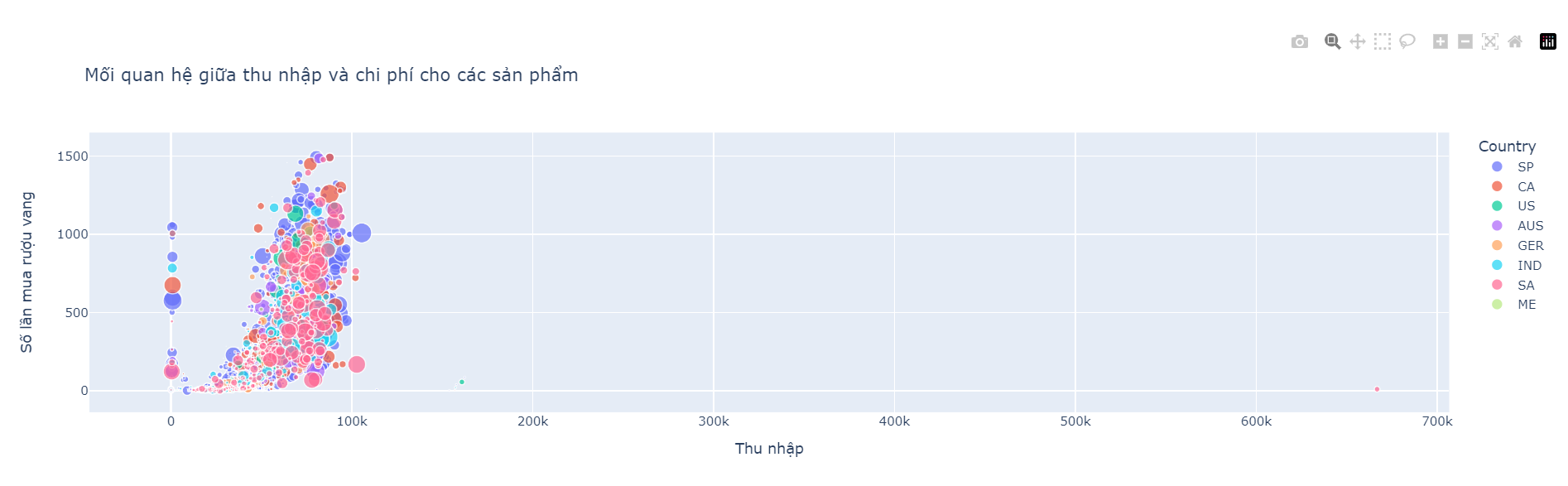




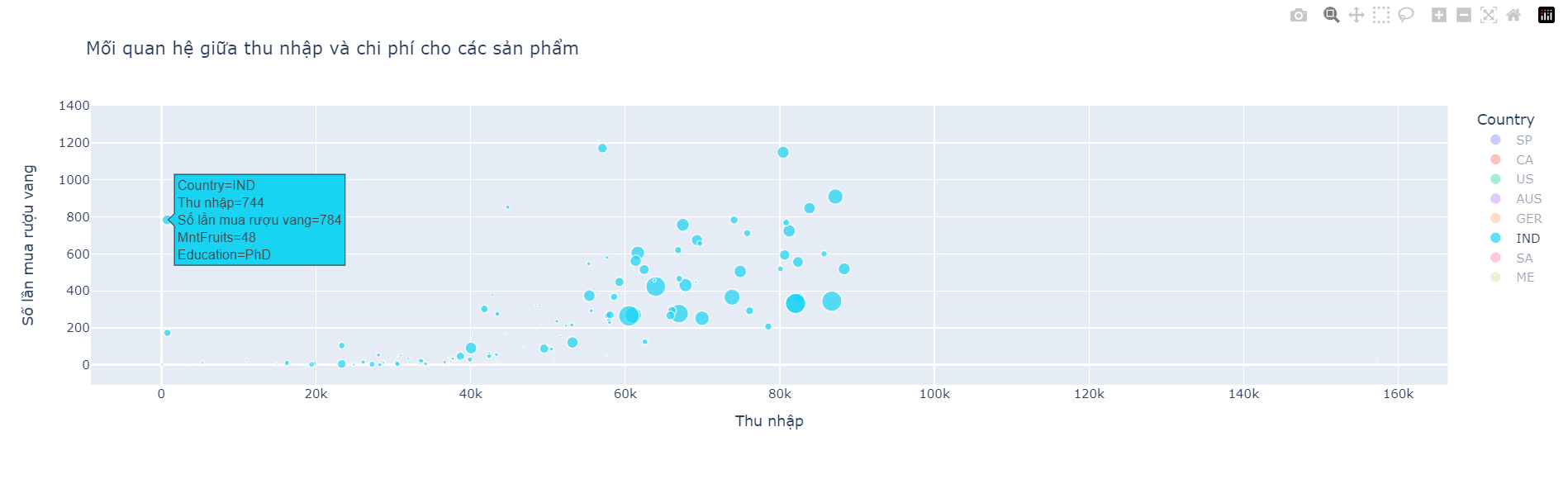


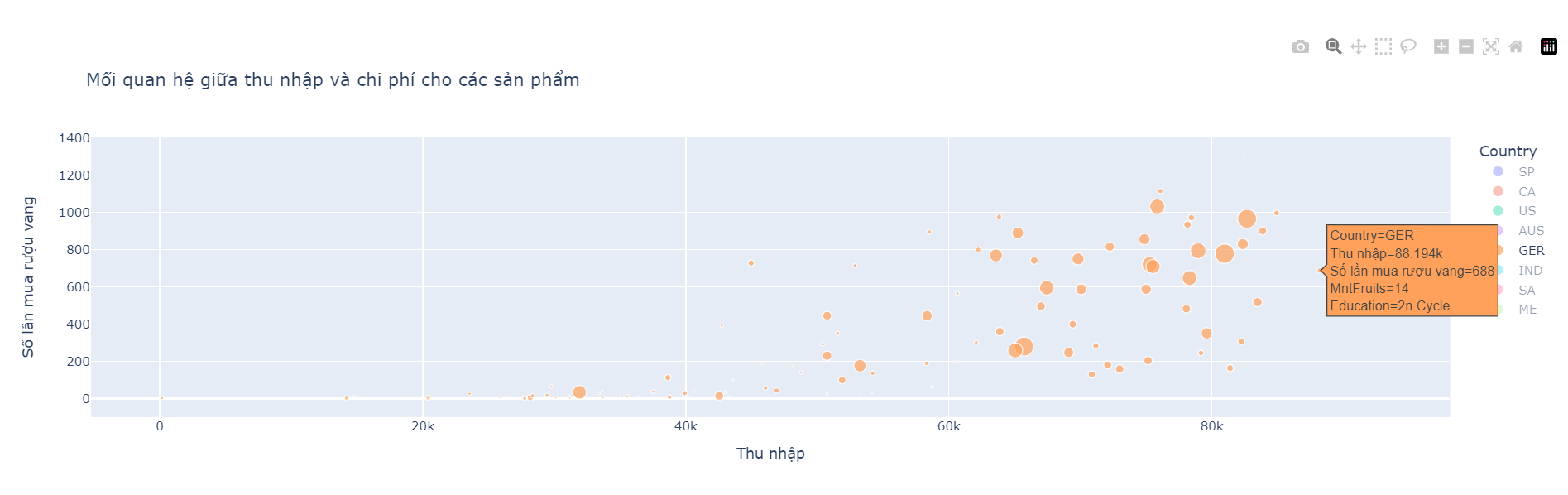
Có thể lựa chọn các sản phẩm nào bạn cần quan tâm để xem tổng số lượng đã mua về món sản phẩm đó

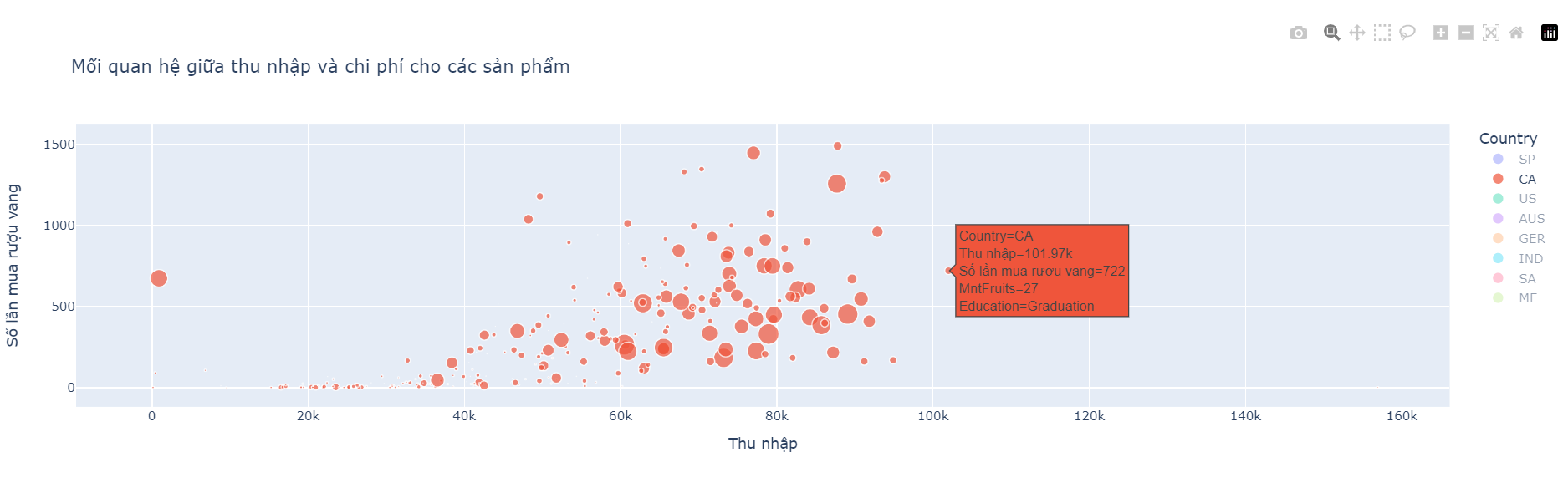
1. **Mối quan hệ giữa thu nhập và chi phí cho các sản phẩm**

****

Biểu đồ scatter này thể hiện mối quan hệ giữa thu nhập của khách hàng với số lần mua sản phẩm rượu vang, với mỗi màu sắc biểu thị mỗi vùng và kích thước điểm biểu thị số lần mua sản phẩm trái cây. Biểu đồ cho thấy rằng có một sự tương quan giữa thu nhập và số lần mua sản phẩm rượu vang, nghĩa là khách hàng có thu nhập cao hơn thường mua sản phẩm rượu vang nhiều hơn. Ngoài ra, biểu đồ cũng cho thấy một sự đa dạng trong mức độ chi tiêu cho sản phẩm trái cây giữa các quốc gia khác nhau.

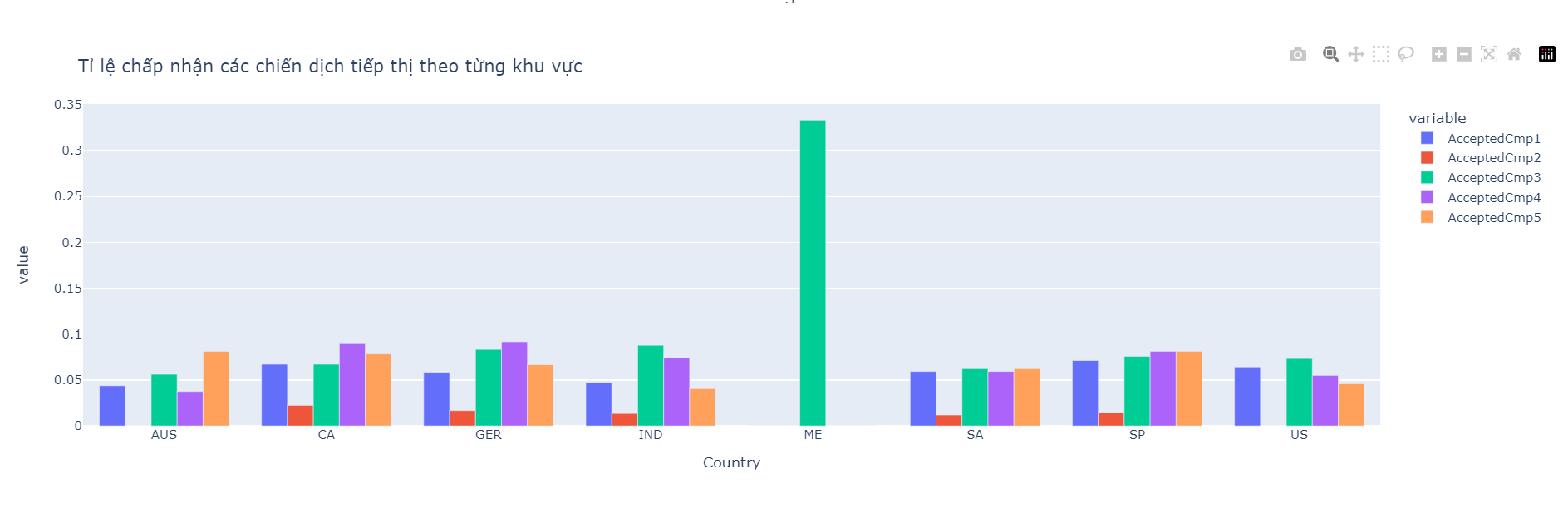






Có thể xem thu nhập của khách hàng với số lần mua rượu vang và trái cây theo từng vùng bằng cách ẩn các vùng khác đi

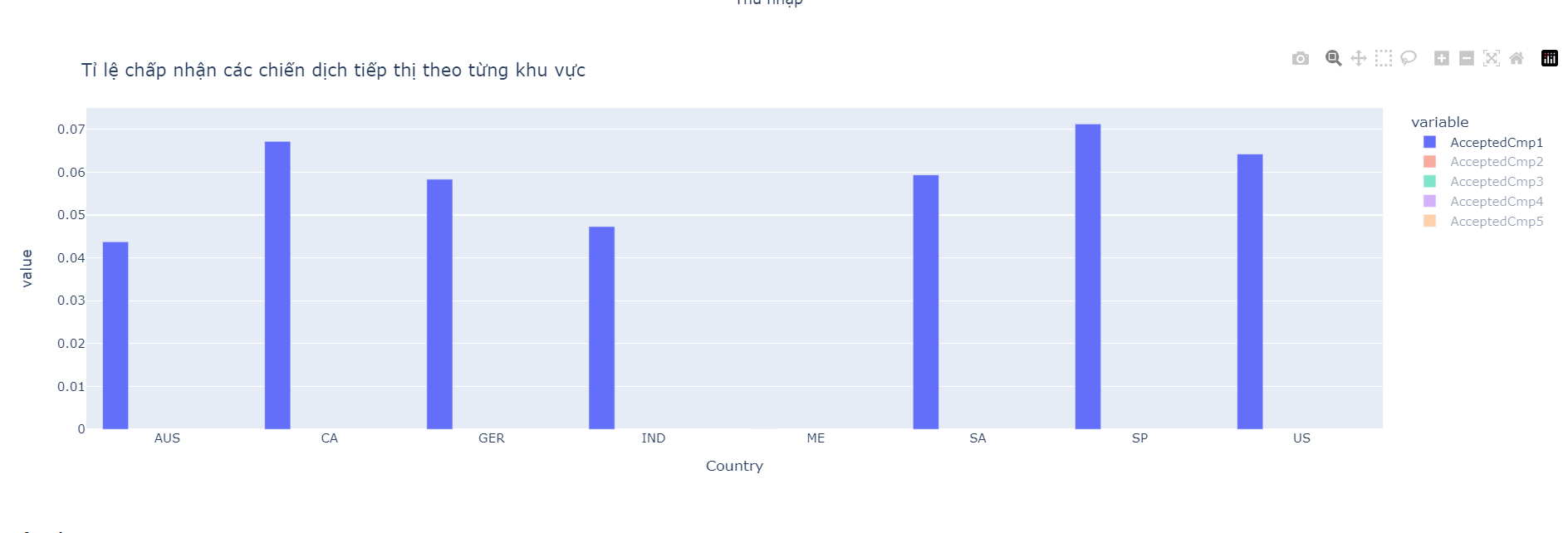
1. **Biểu đồ so sánh tỉ lệ chấp nhận chiến dịch tiếp thị theo vùng**

****

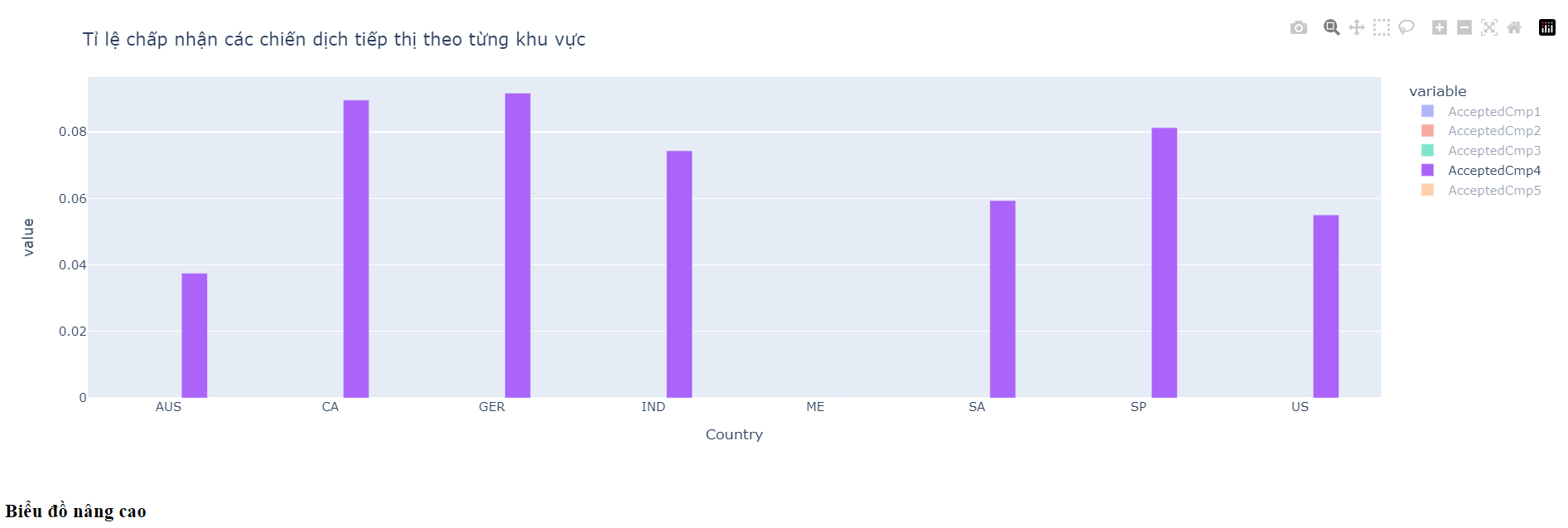
Biểu đồ thể hiện tỉ lệ chấp nhận các chiến dịch tiếp thị của khách hàng theo từng vùng (châu lục) trong dữ liệu. Biểu đồ này sử dụng cột dọc để biểu thị tỉ lệ chấp nhận, và sử dụng cột ngang để biểu thị các chiến dịch tiếp thị khác nhau, bao gồm AcceptedCmp1, AcceptedCmp2, AcceptedCmp3, AcceptedCmp4 và AcceptedCmp5.

Mỗi thanh cột thể hiện tỉ lệ chấp nhận các chiến dịch tiếp thị ở từng vùng khác nhau. Khi di chuột qua các thanh cột, số liệu chi tiết về tỉ lệ chấp nhận của từng chiến dịch tiếp thị sẽ được hiển thị.

Có thể xem từng đợt tiếp thị bằng cách ẩn các đợt tiếp thị khác đi bằng cách click vào phần chú thích ở bên phải biểu đồ







1. **Biểu đồ nâng cao**

**7.1. Thu nhập của khách hàng dựa trên trình độ học vấn**

**Chart, box and whisker chart

Description automatically generated**

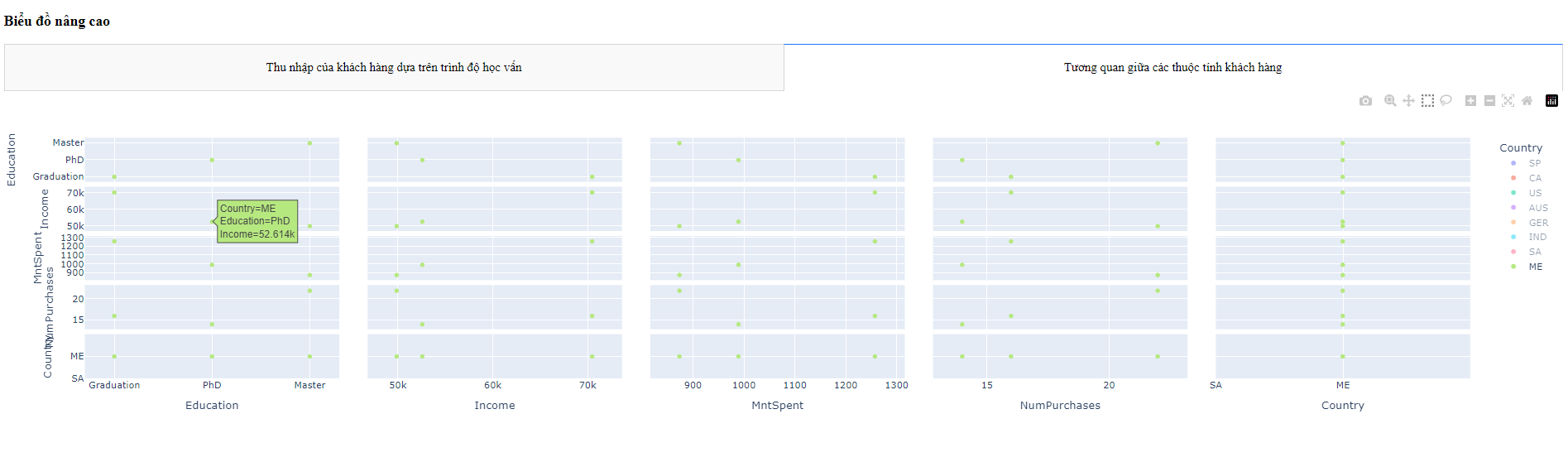
Biểu đồ trên thể hiện được mức thu nhập trung bình và thu nhập lớn nhất dựa trên trình độ học học vấn của khách hàng. Từ đó ta có thể thấy dựa theo trình độ học vấn của khách hàng ta sẽ có những mức thu nhập khác nhau

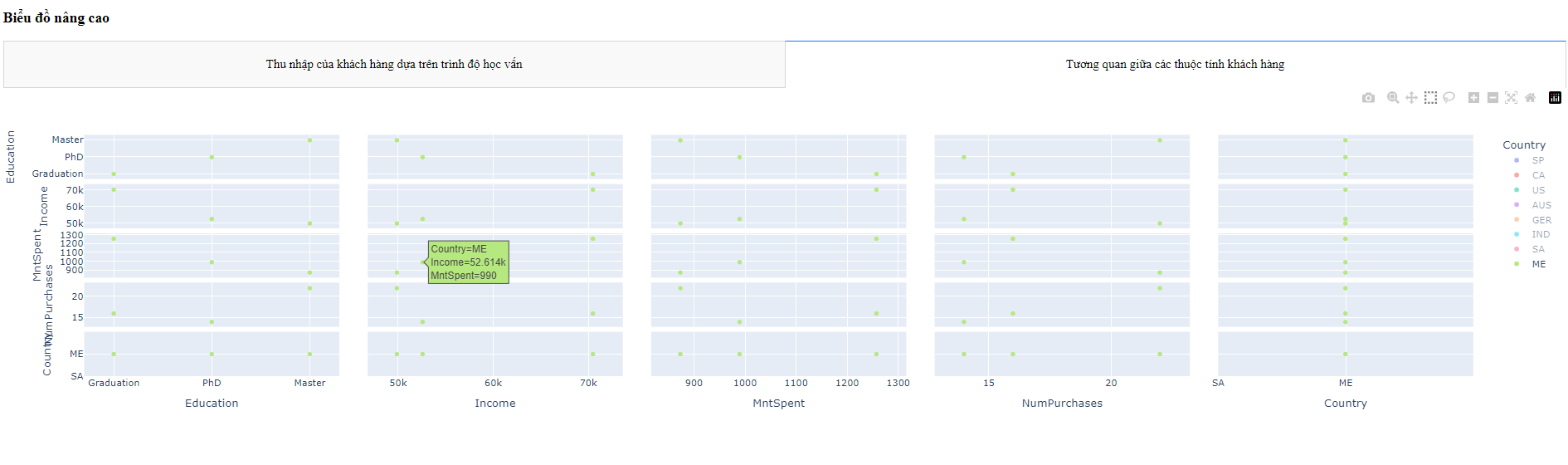
**7.2. Lưới chọn thuộc tính**

****

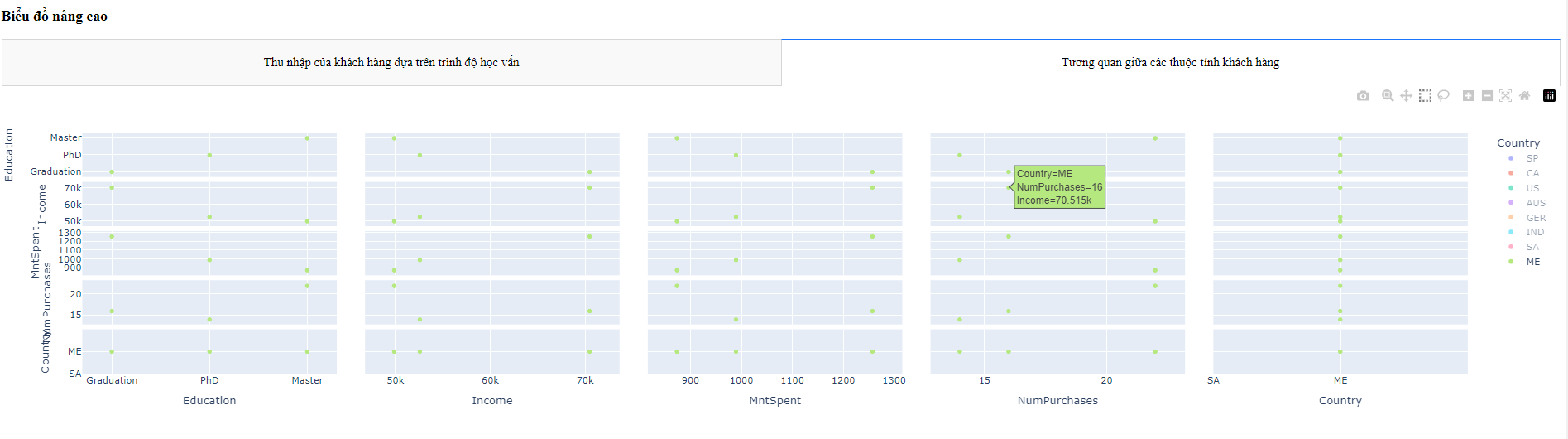
Biều đồ thể hiện đô tương quan giữa thu nhập (Income), tổng số tiền đã chi tiêu cho sản phẩm (MntSpent) của khách hàng, trình độ học vấn (Education), số lượng sản phẩm khách hàng đã mua (NumPurchases) và , được phân loại theo khu vực (Country)

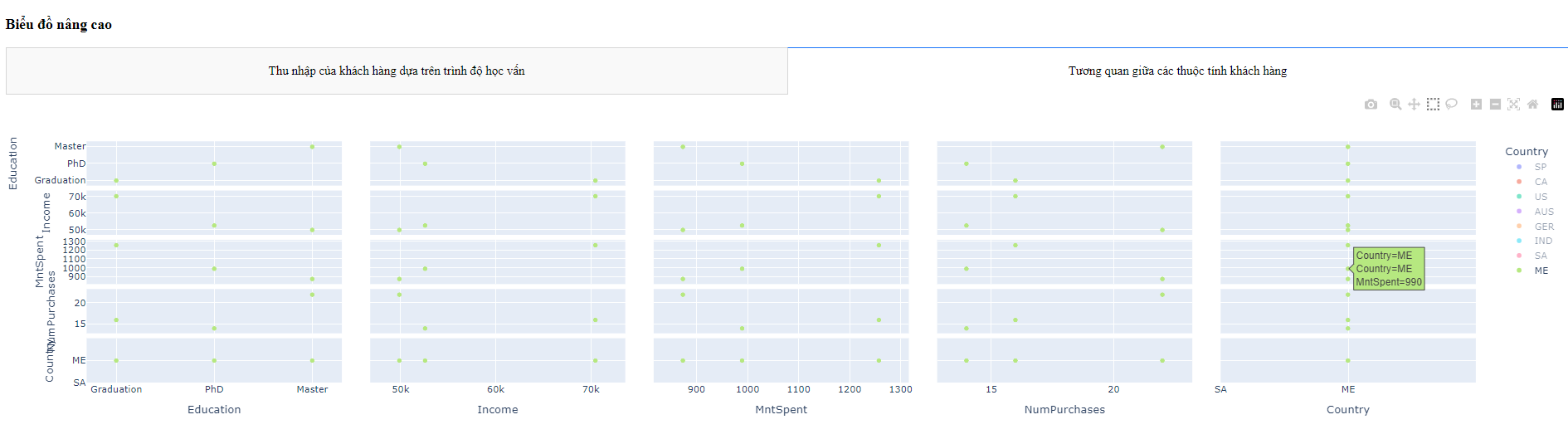
Chúng ta có thể xem mỗi một khu vực bằng cách ẩn các khu vực khác đi bằng cách click vào các khu vực (Country) ở bên góc phải biểu đồ





****

****

****

Dựa vào biểu đồ trên ta có thể thấy khu vực ME có thu nhập (Income), tổng số tiền đã chi tiêu cho sản phẩm (MntSpent) của khách hàng, trình độ học vấn (Education), số lượng sản phẩm khách hàng đã mua (NumPurchases)

Từ đó đưa ra được các kết luận như:

* Khách hàng có thu nhập 52.614k có tổng số tiền chi tiêu cho sản phẩm là 990
* Tổng số tiền chi tiêu cho sản phẩm là 990 thì có số lượng sản phẩm đã mua là 14 …

1. **Đường trung bình thu nhập theo năm sinh**

**Graphical user interface, chart, application

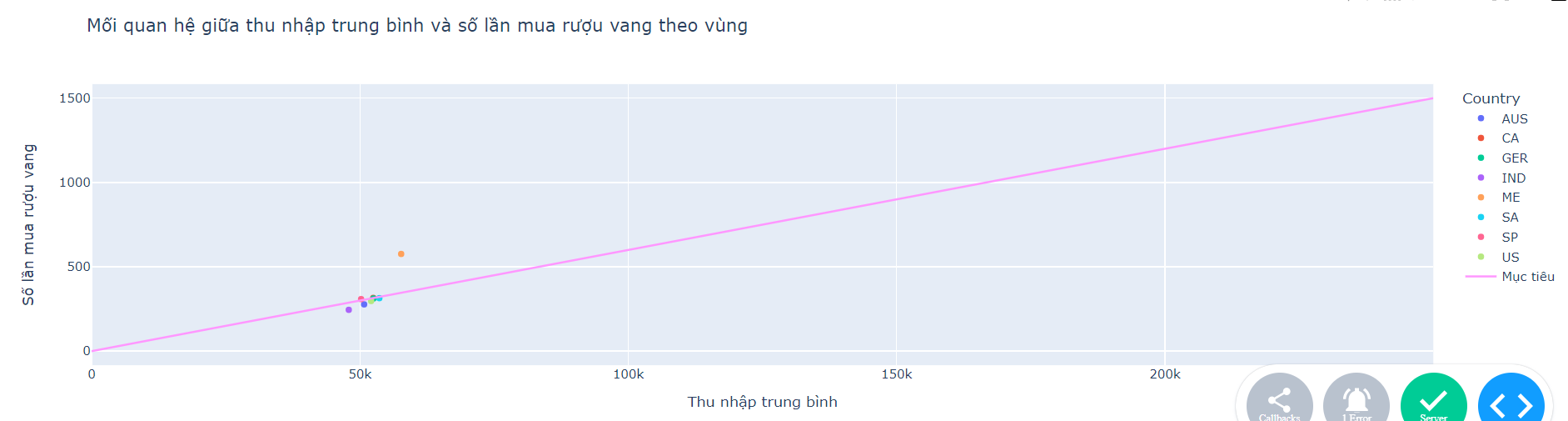
Description automatically generated**

Biểu đồ thể hiện đường (line chart) thể hiện xu hướng thu nhập trung bình của các khách hàng trong tập dữ liệu theo từng năm sinh. Các điểm trên đường biểu thị giá trị trung bình của thu nhập của các khách hàng trong từng nhóm năm sinh, còn đường thẳng nối các điểm biểu thị xu hướng tăng hoặc giảm của thu nhập trung bình theo từng nhóm năm sinh.

Ở đây, dữ liệu được chia thành các nhóm năm sinh, với trục x biểu thị cho các nhóm năm sinh, và trục y biểu thị cho giá trị trung bình của thu nhập. Bên cạnh đó, biểu đồ còn có thêm một số thông tin trên trục x và trục y để giải thích ý nghĩa của các trục này.

Từ biểu đồ, chúng ta có thể nhận thấy rằng các khách hàng sinh vào khoảng từ năm 40 đến 60 có xu hướng có thu nhập cao hơn so với các khách hàng sinh vào khoảng thập niên trước đó hoặc sau đó. Các khách hàng sinh vào khoảng từ năm 70 đến 100 có xu hướng có thu nhập trung bình thấp hơn so với các nhóm năm sinh khác.

1. **Mối quan hệ giữa thu nhập trung bình và số lần mua rượu vang theo vùng và vẽ đường mục tiêu cho các sản phẩm**

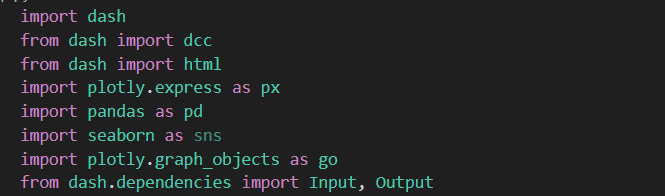


Biểu đồ trên thể hiện mối quan hệ giữa thu nhập trung bình của khách hàng và số lần mua rượu vang (MntWines) của khách hàng theo từng vùng. Mỗi điểm trên biểu đồ đại diện cho một vùng và có màu sắc khác nhau để phân biệt.

Đường thẳng "Mục tiêu" là một đường thẳng chỉ mục tiêu, thể hiện mức độ cần đạt được của số lần mua rượu vang theo thu nhập trung bình. Đường thẳng này có dạng từ điểm (0,0) đến điểm (250000, 1500), tức là nếu thu nhập trung bình của một khách hàng là 250.000 đô la thì mục tiêu là mua 1.500 lần rượu vang.

Từ biểu đồ, ta có thể thấy rằng số lần mua rượu vang tăng theo thu nhập trung bình của khách hàng. Nhìn chung, các quốc gia có thu nhập trung bình cao hơn có xu hướng mua rượu vang nhiều hơn. Tuy nhiên, có một số quốc gia với thu nhập trung bình cao nhưng lại mua rượu vang ít hơn so với những quốc gia khác, và ngược lại.

1. **Các công nghệ/phần mềm sử dụng**

****

* + Dash: dùng để tạo ra dashboard để chạy trên localhost: <http://127.0.0.1:8050/>
  + Pandas: Dùng để xử lí file
  + Plotly: Dùng để vẽ biểu đồ trực quan

1. **Kết luận**

Đa số khách hàng trong tập dữ liệu này đều có thu nhập trung bình ở mức 30.000 - 60.000

Khách hàng trong độ tuổi từ 45 đến 55 có thu nhập trung bình cao nhất, trong khi độ tuổi trẻ hơn thì thu nhập trung bình thấp hơn. Điều này có thể giải thích bằng việc khách hàng ở độ tuổi trung niên thường có nhiều kinh nghiệm trong công việc, nên có thu nhập cao hơn.

Trong số các sản phẩm được khách hàng mua, rượu vang là sản phẩm được mua nhiều nhất, tiếp đó là các loại thịt, hải sản, trái cây và các sản phẩm vàng. Điều này cho thấy sở thích và nhu cầu của khách hàng trong tập dữ liệu này

Cuối cùng, khách hàng có thu nhập trung bình cao hơn thường mua nhiều sản phẩm hơn, đặc biệt là các sản phẩm cao cấp như vàng. Điều này cho thấy khách hàng có thể có xu hướng mua sắm dựa trên thu nhập của mình.

Các sản phẩm thành công nhất là rượu vang và thịt (tức là khách hàng trung bình đã chi tiêu nhiều nhất cho các mặt hàng này) .Đề xuất: tập trung các chiến dịch quảng cáo để tăng doanh số của các mặt hàng ít phổ biến hơn).

Chiến dịch 2 là chiến dịch kém hiệu quả ở mọi khu vực được áp dụng. Có thể loại bỏ chiến dịch 2 hoặc áp dụng các mô hình của chiến dịch 3,4 sang chiến dịch 2

Khuyến khích tập trung quảng cáo các chiến dịch trong các phương thức bán hàng thành công hơn để tiếp cận nhiều khách hàng hơn và cũng để tăng doanh số của các sản phẩm hoạt động kém.

**TÀI LIỆU THAM KHẢO**

[1] Yassinesfaihi (2023) Uncovering customer insights - eda [plotly], Kaggle. Kaggle. Available at: <https://www.kaggle.com/code/yassinesfaihi/uncovering-customer-insights-eda-plotly> (Accessed: May 6, 2023).

[2] YuehHanChen (2021) Marketing\_analytics/marketing\_data.CSV at main·yuehhanchen/marketing\_analytics,GitHub.Available at: <https://github.com/YuehHanChen/Marketing_Analytics/blob/main/marketing_data.csv> (Accessed: May 6, 2023).

[3] Hamburger22 (2021) *Market\_data\_analysis*, *Kaggle*. Kaggle. Available at: <https://www.kaggle.com/code/hamburger22/market-data-analysis/notebook> (Accessed: May 6, 2023).