**TRƯỜNG ĐẠI HỌC KINH TẾ - ĐẠI HỌC ĐÀ NẴNG**

**CUỘC THI DATAGOTTALENT**

**BÁO CÁO BÀI THI**

**ĐỀ BÀI PHÂN TÍCH BÁN HÀNG LAPTOP**

Nhóm :

Thành viên:

**Nguyễn**

*Đà Nẵng, ngày 23 tháng 01 năm 2022*

**MỤC LỤC**

[MỞ ĐẦU 1](#_Toc95600126)

[I. TIỀN XỬ LÝ DỮ LIỆU 4](#_Toc95600127)

[***1.*** ***Xử lý nhiễu lỗi, mất mát dữ liệu*** 4](#_Toc95600128)

[***2.*** ***Loại bỏ thuộc tính dư thừa và merge table*** 4](#_Toc95600129)

[II. MỤC TIÊU PHÂN TÍCH 4](#_Toc95600130)

[**1.** **Goal** 4](#_Toc95600131)

[**2.** **Character** 5](#_Toc95600132)

[**3.** **Sự liên kết và tương tác của các đối tượng** 5](#_Toc95600133)

[**4.** **Người nghe** 5](#_Toc95600134)

[III. TRÌNH BÀY PHÂN TÍCH 7](#_Toc95600135)

[**Câu 1: Chân dung tập khách hàng** 7](#_Toc95600136)

[a. Tổng quan về khách hàng 7](#_Toc95600137)

[b. Vòng đời khách hàng 9](#_Toc95600138)

[c. Chu kì mua hàng 9](#_Toc95600139)

[d. Phân nhóm khách hàng 10](#_Toc95600140)

[**Câu 2: Tình hình tiêu thụ máy tính và bán hàng của các cửa hàng** 11](#_Toc95600141)

[a. Giá trung bình theo cửa hàng 12](#_Toc95600142)

[b. Các cửa hàng có khuyến mãi (Promotion) 13](#_Toc95600143)

[c. Doanh thu theo từng cửa hàng và theo từng kênh 14](#_Toc95600144)

[d. Số lượng máy bán ra theo thương hiệu 15](#_Toc95600145)

[e. Số lượng bán theo thời gian 16](#_Toc95600146)

[f. Top 10 người mua hàng nhiều nhất 16](#_Toc95600147)

[g. Yếu tố khách hàng quan tâm tới sản phẩm 17](#_Toc95600148)

[**Câu 4: Dưới góc nhìn của nhà phân tích dữ liệu, bạn có tư vấn gì cho một sàn giao dịch điện tử về mảng máy tính** 17](#_Toc95600149)

[*h.* Một mô hình học máy đơn giản 18](#_Toc95600150)

[i. Đánh giá mô hình 19](#_Toc95600151)

[j. Cải tiến mô hình 19](#_Toc95600152)

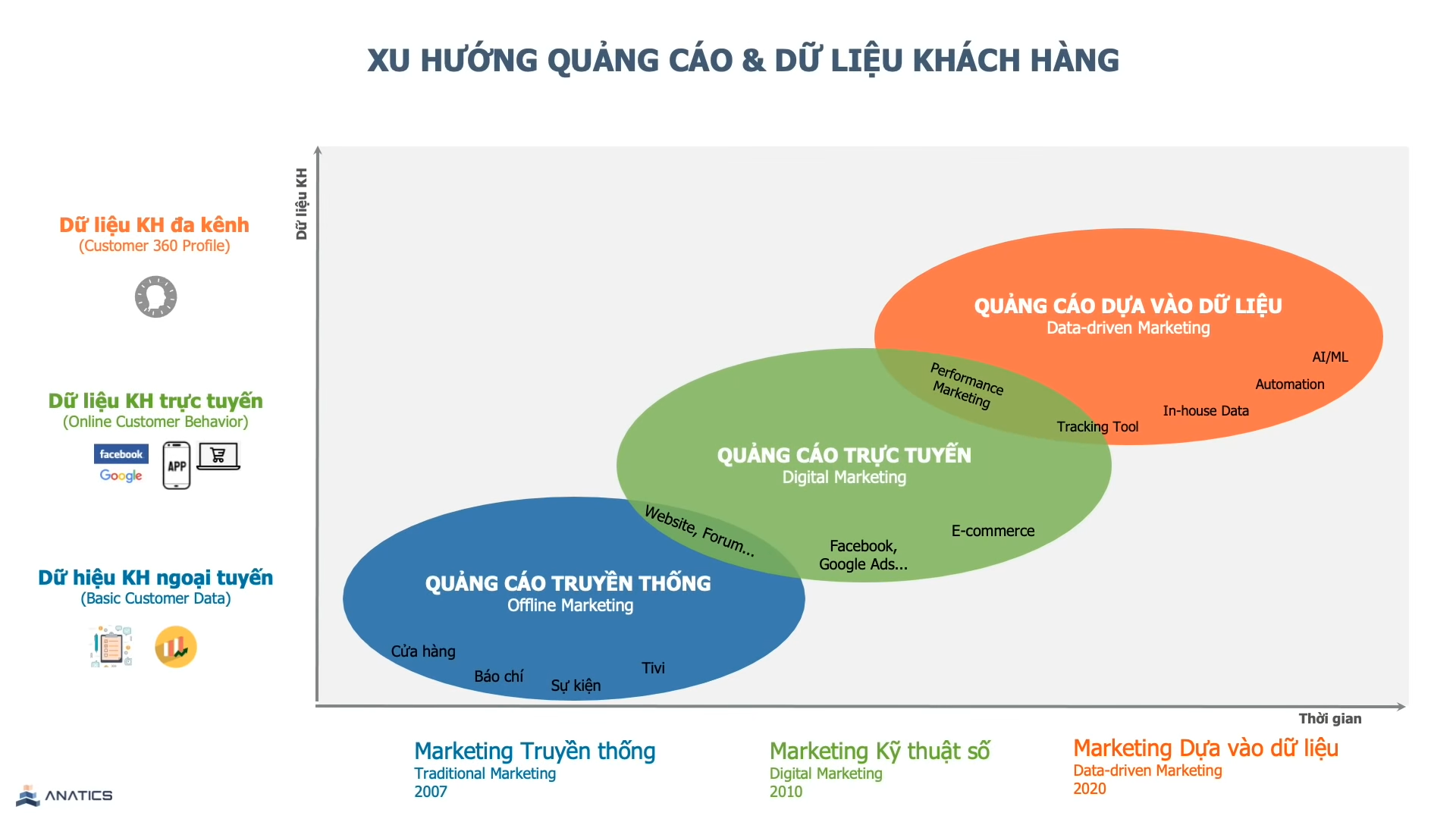
[KẾT LUẬN 20](#_Toc95600153)

# **MỞ ĐẦU**

Trong thời kì đại dịch cùng với sự phát triển của công nghệ thông tin và truyền thông, dữ liệu đặc biệt là một phần quan trọng, dữ liệu đã được các doanh nghiệp, công ty công nghệ, nhà nước, … thu thập cách đây nhiều thời gian, với lượng dữ liệu lớn như vậy làm sao xử lý phân tích và đưa ra những quyết định để nâng cao an ninh quốc gia và tăng hiệu quả sản xuất, hay cụ thể đối với doanh nghiệp là tăng doanh số doanh thu bán hàng, xây dựng chân dung khách hàng, giữ chân khách hàng, … là những câu hỏi khó được đặt ra cho những nhà phân tích dữ liệu, khoa học dữ liệu, kĩ sư dữ liệu, kĩ sư học máy xử lý.

Để hiểu thêm xu hướng sử dụng dữ liệu trong doanh nghiệp với hoạt động kinh doanh, ta xem xét dưới đây là “xu hướng quảng cáo và dữ liệu khách hàng” phát triển theo từng giai đoạn:

* Marketing truyền thống: nơi ta chỉ có thể thu thập email hay số điện thoại khách hàng
* Marketing kỹ thuật số: khi mà internet phát triển hơn xuất hiện các website, platform có thể tiếp cận khách hàng có thể thu thập nhiều thông tin khách hàng hơn từ đa kênh
* Marketing dựa vào dữ liệu: các doanh nghiệp xây dựng các hệ thống phân tích tổng hợp và kết hợp của đội ngũ kĩ sư phân tích xử lý dữ liệu



Hình 0–1: Source Anatics

Từ việc có được thông tin khách hàng từ đa kênh chúng ta có thể phân tích dựa trên số liệu để đưa ra các quyết định đúng đắn cho doanh nghiệp nhằm đạt được hiệu quả bán hàng, giữ chân khách hàng, giải quyết các bài toán cross-sale, up-sale, …

Cuộc thi Datagottalent mô phỏng việc phân tích dữ liệu trong doanh nghiệp giúp sinh viên có thể nắm bắt tình hình thực tế và vận dụng kiến thức để giải quyết các yêu cầu thách thức đặt ra. Nhận được bộ dữ liệu cùng với các câu hỏi gợi ý phân tích: (1) Mô tả đặc trưng của khách hàng mua máy tính, (2) Bạn có nhận xét gì về tình hình tiêu thụ máy tính và bán hàng của các cửa hàng, (3) Bạn sẽ tư vấn cho người mua như thế nào để đạt hiệu quả bán hàng cao nhất, (4) Dưới góc nhìn của nhà phân tích dữ liệu, bạn có tư vấn gì cho một sàn giao dịch điện tử mảng máy tính. Chúng tôi trình bày bài phân tích của mình thành 3 phần chính (I) Tiền xử lý dữ liệu, (II) Mục tiêu phân tích, (III) Trình bày phân tích và phần Kết luận. Phần (I) Chúng tôi sẽ trình bày cách xử lý dữ liệu, nhiễu lỗi, dữ liệu dư thừa, tổng quan dữ liệu. Phần (II) Mục tiêu phân tích, sẽ được trình bày về đối tượng, mục đích cuối cùng, người nghe phân tích. Phần (III) chúng tôi lồng ghép câu 1,2 ở trên cùng lúc đưa ra nhận xét để giải quyết vấn đề ở câu 3, câu 4 chúng tôi sẽ xây dựng một mô hình học máy để giải quyết bài toán dự đoán gợi ý mua hàng cho khách hàng. Phần kết luận là phần tổng hợp nhận xét của bài phân tích.

Trong báo cáo này, chúng tôi trình bày một số phân tích đánh giá về tình hình tiêu thụ laptop cũng như chân dung khách hàng các nhận xét và áp dụng mô hình học máy trên phương diện tìm tòi học hỏi nâng cao kĩ năng tuy nhiên vẫn còn rất nhiều điểm chưa thể hoàn thiện.

# **TIỀN XỬ LÝ DỮ LIỆU**

Dataset: là bộ dữ liệu khoảng 1k-2k quan sát ở mỗi 2 bảng customer\_profile và sale, dữ liệu sinh viên với hành vi mua laptop trên 1 sàn thương mại điện tử, trong phạm vi 1 trường ĐH trong 1 năm (là năm 2020).

## ***Xử lý nhiễu lỗi, mất mát dữ liệu***

Ở đây chúng tôi tóm tắt là cách xử lý dữ liệu dư thừa mất mát cụ thể hơn ghi ở file notebook.

|  |  |
| --- | --- |
| Trường xử lý | Cách xử lý |
| CustomerID | Drop row |
| OrderID | Drop col |
| Channel | Mất mát data điền all online |
| PaymentType | Điền tất cả là thẻ |
| Brand | Điền nhãn hiệu bán chạy nhất Asus |
| Shop | Điền theo shop bán nhiều hàng nhất |
| SalePrice | Điền theo giá trung bình sản phẩm |
| Reference | Điền theo giá trị thuộc tính xuất hiện nhiều nhất |

## ***Loại bỏ thuộc tính dư thừa và merge table***

Để tiện cho việc phân tích và xây dựng mô hình học máy chúng tôi quyết định gộp 2 bảng CustomerProfile và Sale, chúng tôi loại bỏ thuộc tính OrderID vì thấy rằng rất khó để xác định giao dịch theo thuộc tính này (vì xuất hiện trùng lặp OrderID ở các cửa hàng khác nhau – thứ mà là thuộc tính định danh, và giao dịch mua laptop là giao dịch đơn chiếc)

# **MỤC TIÊU PHÂN TÍCH**

## **Goal**

* Mô tả các đặc trưng của khách hàng mua máy tính, sàn thương mại
* Hiểu được tình hình tiêu thụ máy tính và bán hàng của các cửa hàng
* Thông qua dữ liệu sẽ tư vấn cho người mua hiệu quả đạt kết quả bán hàng cao nhất.
* Tăng hiệu quả đáng kể cho các chiến dịch marketing để tiếp cận tới khách hàng.

Vì thông tin sale gồm cả thông tin các shop, brand nhưng chúng tôi sẽ tập tập trung phân tích giúp các nhà quản lý sàn thương mại điện tử kinh doanh hiệu quả.

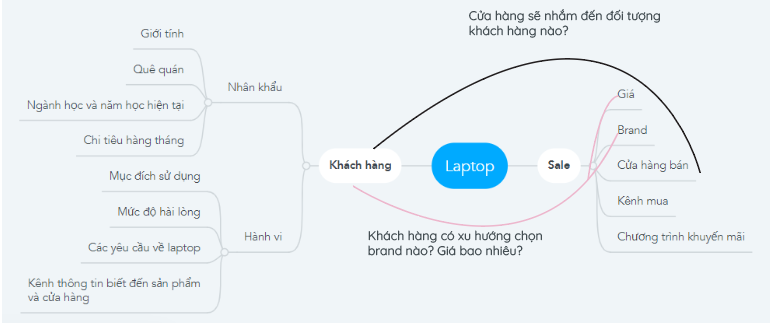
Các bài toán upsale, cross-sale quen thuộc với các doanh nghiệp sẽ không được phân tích vì không đủ dữ liệu, dữ liệu đơn thuần là bán laptop, thay vào đó chúng tôi tập trung mô tả chân dung khách hàng, tìm khách hàng mới, quảng cáo hiệu quả.

## **Character**

Tập trung vào phân tích khách hàng mua laptop nên sẽ xoay quanh 2 đối tượng là khách hàng và Sale (Bán hàng trên sàn thương mại điện tử)

* Khách hàng:
* Các thông tin về nhân khẩu học bao gồm: Giới tính, quê quán, nghề nghiệp gia đình, chi tiêu hàng tháng và chuyên ngành hiện tại đang học.
* Các thông tin về hành vi của khách hàng: Thời gian sử dụng, mục đích sử dụng, mức độ hài lòng, các yếu tố yêu cầu
* Sale: Thương hiệu máy, giá bán, cửa hàng, kênh, chương trình khuyến mãi

## **Sự liên kết và tương tác của các đối tượng**



## **Người nghe**

* Người nghe (Who): Trưởng bộ phận bán hàng và bộ phận Marketing,
* Nghe ở đâu (Where): Cuộc họp kinh doanh quý 1 năm 2022
* Nghe về cái gì (What): Phân tích hành vi của khách hàng khi mua máy tính
* Nghe để làm gì (Why):
* Đối với bộ phận bán hàng: Dựa vào các thông tin của khách hàng sẽ có thể xác định được nhu cầu để có thể nhập hàng phù hợp
* Đối với bộ phận Marketing: Dựa vào đó để lên các chiến dịch quảng cáo

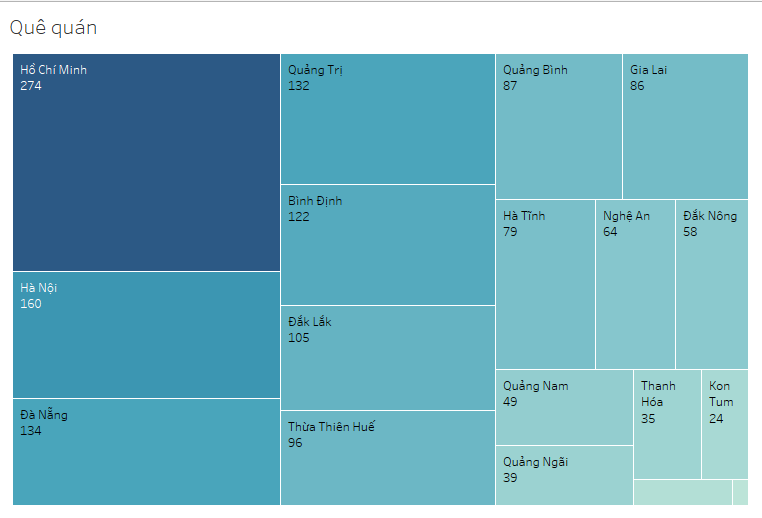
=> Giúp công ty tối đa hóa lợi nhuận.

# **TRÌNH BÀY PHÂN TÍCH**

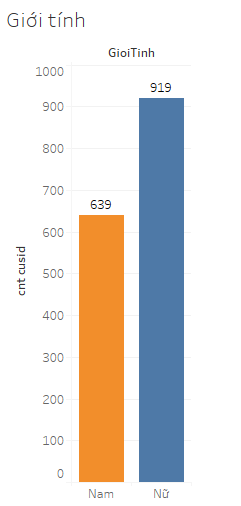
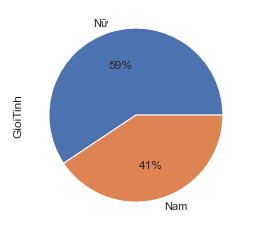
## **Câu 1: Chân dung tập khách hàng**

Dưới đây là một số thông tin cơ bản về chân dung khách hàng mà chúng tôi thấy quan trọng.

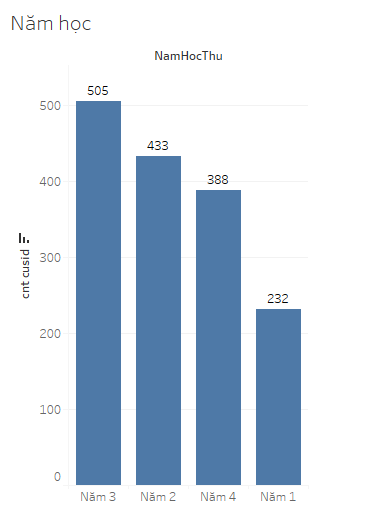
1. Tổng quan về khách hàng



Khách hàng mua lap top phân bổ ở các thành phố lớn, và đặc biệt 3 thành phố lớn: Hồ Chí Minh, Hà Nội, Đà Nẵng



Tỷ lệ nữ nhiều hơn nam => Sinh viên trường kinh tế.



Sinh viên năm 3 là lượng khách hàng lớn nhất (32,4%) do đây là năm bước vào chuyên ngành (và có thể sinh viên đi làm, thực tập) nên việc có laptop là một yêu cầu bắt buộc.

Năm 1 có số lượng khách ít nhất do mới vào trường thì chủ yếu học các môn đại cương, chưa thực sự cần thiết đến laptop.

Tuy nhiên với phân bố này ta vẫn tập trung vào cả học sinh của ở 4 năm học.

1. Vòng đời khách hàng

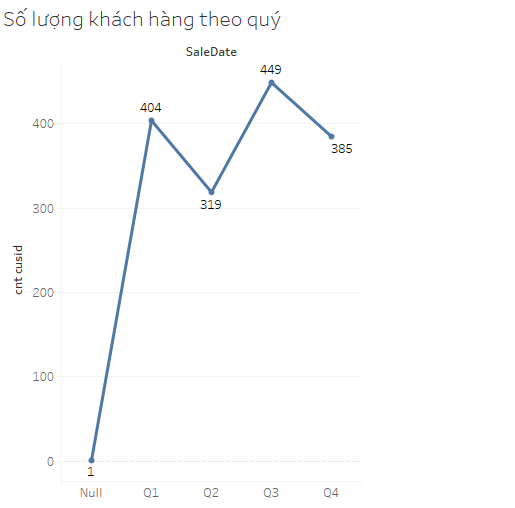
Số lượng khách hàng mua laptop trung bình rơi vào 2-3 chiếc trên thực tế, theo bộ dữ liệu mua nhiều nhất 5 và hầu hết 90% mua 1 chiếc để phục phụ nhu cầu sử dụng.



Bảng số lượng laptop 1 khách hàng mua: mua 1 laptop 723, mua 2 laptop 310, mua 3 laptop 54 người, 4: laptop 12 người, 5 laptop 2 người.

Từ vòng đời khách hàng chúng tôi đưa ra gợi ý cho các bên bán hàng trên sàn thương mại tập trung vào việc chăm sóc khách hàng lần đầu tiên mua hàng, và xem xét kết hợp cùng họ quảng cáo cho người thân.

1. Chu kì mua hàng

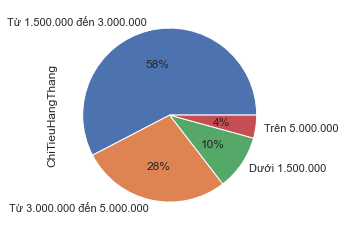


Số lượng khách hàng mua lớn vào quý 1 và quý 3, đây là 2 quý bắt đầu vào kỳ học mới và tạo thành chu kỳ hàng năm.

* Có thể đẩy mạnh chiến dịch quảng cáo vào đầu mỗi kì học để tăng doanh số bán hàng.

1. Phân nhóm khách hàng

Có thể phân nhóm khách hàng theo nhiều cách khác nhau: như theo giới tính, theo nghề nghiệp gia đình, .. Nhưng chúng tôi nhận thấy việc phân loại khách hàng quan trọng nhất là quan tâm yếu tố chi tiêu hàng tháng có thể diễn tả các thông tin khác và phân loại phù hợp cho mục đính kinh doanh mà sàn thương mại hướng tới.



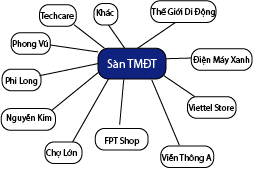
4 Nhóm khách hàng có chi tiêu hàng tháng:

* Từ 1.500.000 đến 3.000.000 634
* Từ 3.000.000 đến 5.000.000 306
* Dưới 1.500.000 113
* Trên 5.000.000 47

Từ đây chúng tôi có thể tập chung vào nhóm đối tượng khách hàng có chi tiêu **Từ 1.500.000 đến 3.000.000** để quan sát phân tích

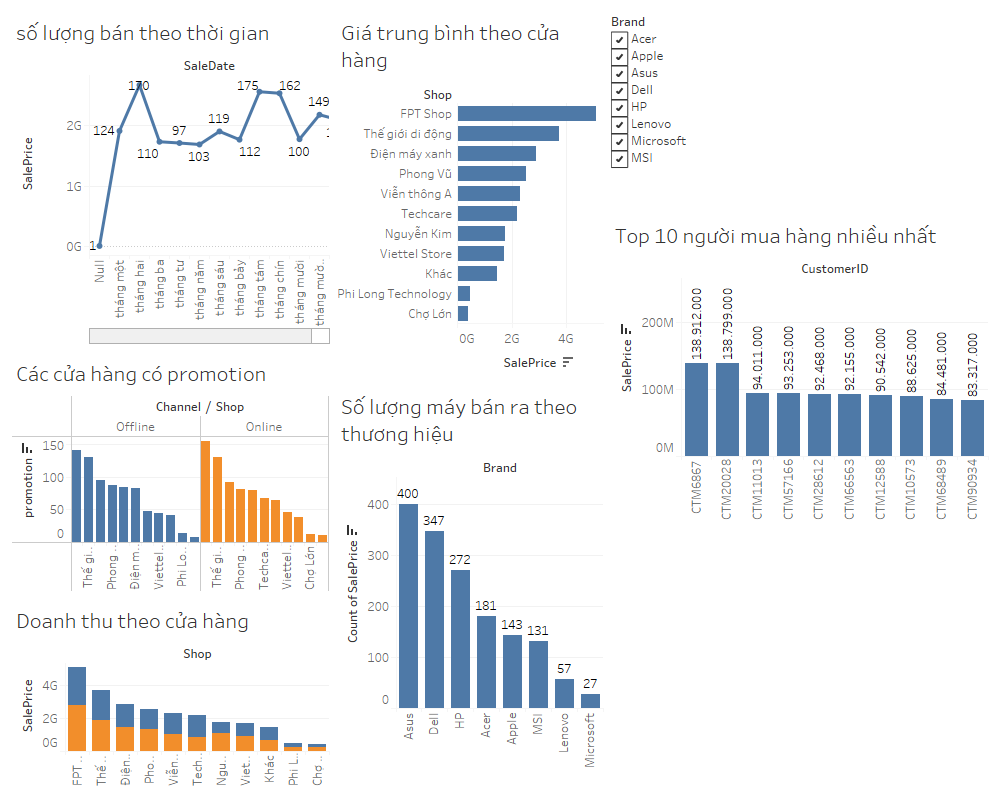
## **Câu 2: Tình hình tiêu thụ máy tính và bán hàng của các cửa hàng**

***Các shop có trên sàn thương mại***

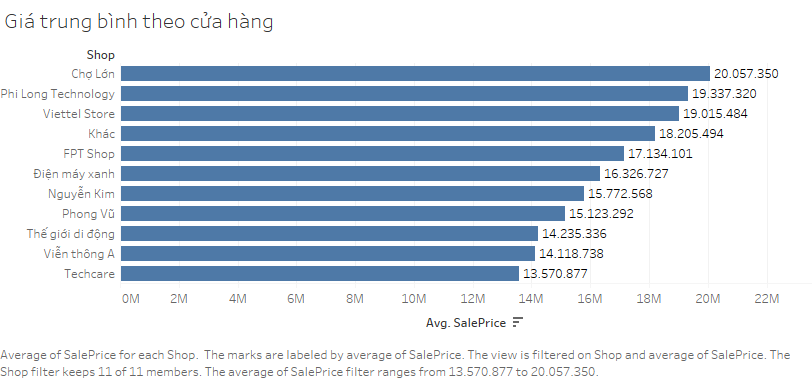


***Dashboard về thông tin bán hàng***

* Để hiểu về tình hình tiêu thụ máy tính và bán hàng của các cửa hàng thì ta quan sát Dashboard dưới.
* Nhằm lên chiến dịch bán hàng dựa vào thời gian khách hàng mua hàng, promotion, doanh thu, số lượng máy bán ra theo thương hiệu.



* 1. Giá trung bình theo cửa hàng



Nội dung chính của biểu đồ trên là giá cả trung bình của sản phẩm theo của hàng.

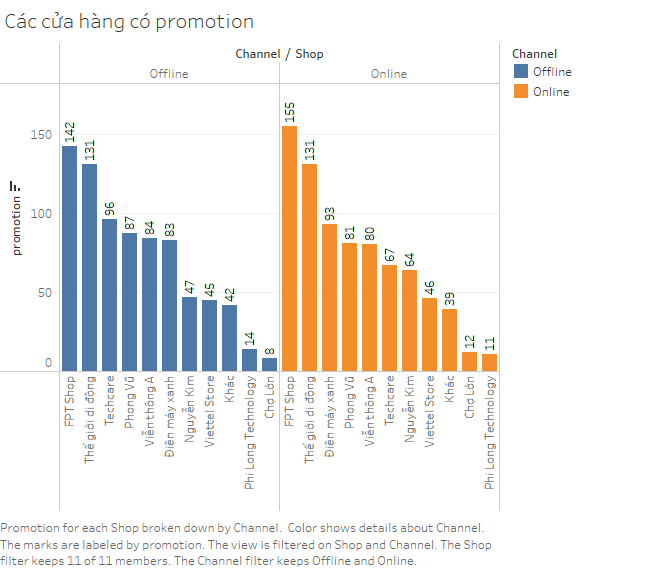
Dạng biểu đồ: Biểu đồ cột.

Qua biểu đồ, nhận định:

* Các shop Chợ Lớn, Phi Long, Viettel Store thì các sản phẩm có mức giá ở phân khúc cao tầm 19-20 triệu.
* Techcare thì cung cấp các sản phẩm laptop tầm thấp 13,5 triệu.

=> Ta có thể phân phân khúc của cửa hàng. Giúp bộ phận bán hàng sàn TM (Thương mại) biết được hiện nay những cửa hàng nào thuộc phân khúc mức giá nào để có thể điều phối các sản phẩm. Bộ phận marketing có thể lên chiến dịch phù hợp với từng phân khúc của cửa hàng.

* 1. Các cửa hàng có khuyến mãi (Promotion)

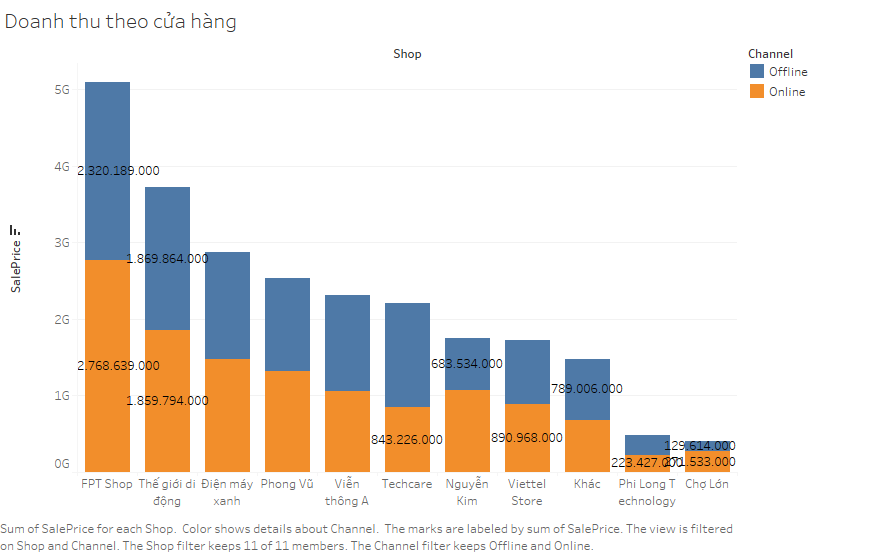


Nội dung chính của biểu đồ trên là xem những cửa hàng nào triển khai chương trình khuyến mãi.

Dạng biểu đồ: Biểu đồ cột.

Qua biểu đồ thì nhận thấy rằng:

* 2 hệ thống lớn là FPT shop và thế giới di động thực hiện nhiều chiến dịch promotion trên cả kênh online và kênh offline. Đối với FPT có 142 lần khuyến mãi trên kênh offline và 155 lần khuyến mãi trên kênh online.
* Chợ Lớn và Phi Long ít áp dụng promotion, tuy nhiên Phi Long tập trung promotion ở kênh offline và chợ lớn chủ yếu ở kênh online.
* Ở đây ta đưa vài nhận định: FPT chính vì khai thác bán hàng dựa trên khuyến mãi cũng nhờ đó doanh thu cao hơn (cao nhất theo số liệu bên dưới) và đồng thời họ cũng phải đẩy giá laptop lên cao hơn
* Đây là một chiến dịch mà đánh vào tâm lý thích mua hàng khuyến mãi của khách hàng được áp dụng ở một hệ thống công ty lớn
  1. Doanh thu theo từng cửa hàng và theo từng kênh

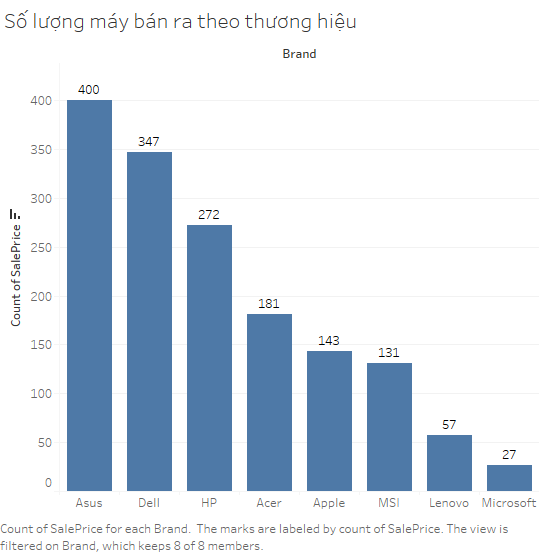


Nội dung chính của biểu đồ trên là doanh thu của từng cửa hàng theo từng kênh.

Dạng biểu đồ: Biểu đồ cột chồng.

Qua biểu đồ thì nhận thấy rằng:

* Hầu hết doanh thu từ 2 kênh online và offline tương đương nhau, tuy nhiên những cửa hàng có giá tầm trung thì có lượng doanh thu từ kênh offline nhiều hơn.
* Cửa hàng FPT Shop có doanh thu nhiều nhất với số tiền là 5.029.028.000 đồng.
* Thế giới di động đứng thứ 2 với doanh thu 3.729.658.000 đồng.
* Cửa hàng Chợ Lớn có doanh thu thấp nhất với doanh thu 401.147.000 đồng.
* Đưa ra khuyến cáo cho doanh nghiệp muốn phát triển kinh doanh cần tập trung vào mảng ofiline đối với những shop có quy mô nhỏ
  1. Số lượng máy bán ra theo thương hiệu

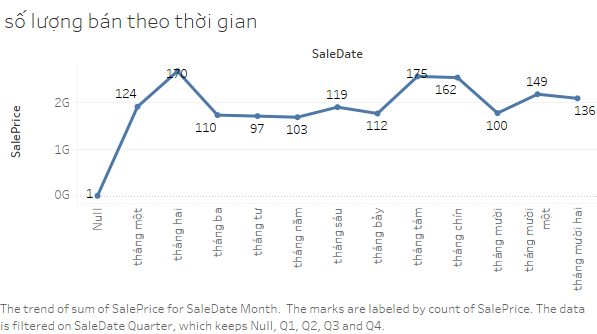


Nội dung chính của biểu đồ trên cho biết số lượng máy bán ra theo từng brand.

Dạng biểu đồ: Biểu đồ cột.

Qua biểu đồ thì nhận thấy rằng:

* Sinh viên kinh tế chủ yếu mua Asus và Dell.
* Trong năm 2020 thì số máy Asus được bán ra là 400 máy, Dell là 347 máy.
* Yếu tố thiết kế đẹp ở Asus và độ bền cũng như thương hiệu ở Dell phù hợp với sinh viên kinh tế. Các doanh nghiệp hay sàn có thể cung cấp thông tin này để chủ doanh nghiệp nhà kinh doanh có thể nhập hàng và quảng cáo nhiều về mặt hàng Dell hay Asus.
  1. Số lượng bán theo thời gian



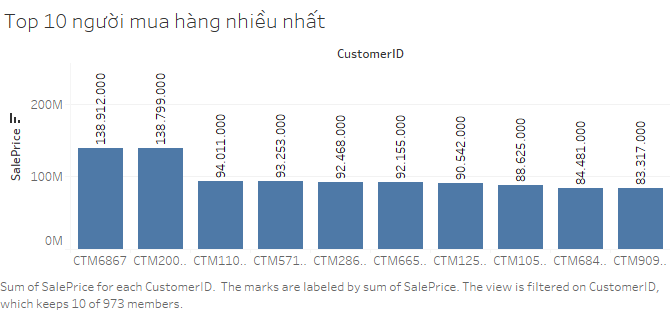
Nội dung chính của biểu đồ trên cho biết số lượng máy bán ra theo thời gian.

Dạng biểu đồ: Biểu đồ đường.

Qua biểu đồ thì thấy rằng, tháng 2-8-9 là 3 tháng có số lượng bán máy nhiều nhất. => Đây là thời điểm bắt đầu học nên nhu cầu sử dụng nhiều hơn.

Như đã phân tích ở trên nhà kinh doanh nên đẩy mạnh quảng cáo và thúc đẩy marketing ở thời gian này.

* 1. Top 10 người mua hàng nhiều nhất

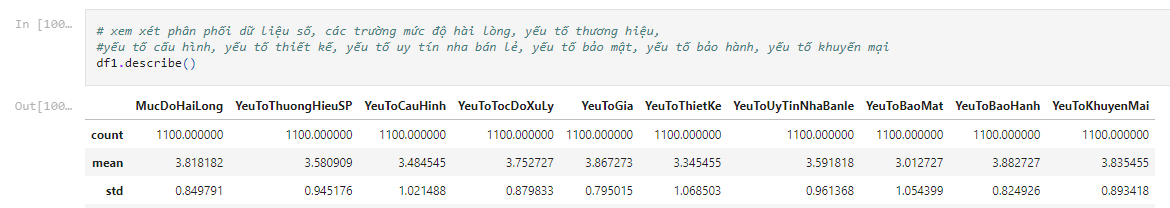


Nội dung chính của biểu đồ trên cho biết top 10 khách hàng mua nhiều nhất.

Dạng biểu đồ: Biểu đồ cột.

Qua biểu đồ nhận thấy rằng:

* Khách hàng có mã CTM20028 có số lượng mua lớn nhất. 9 người còn lại thì số tiền bỏ ra khi mua máy tính tương đương nhau.
* Theo đã phân tích thì số lượng khách hàng mua lớn nhất là 5 máy laptop.Và trung bình 1 người mua hàng có vòng đời mua sản phẩm laptop là 2-3 chiếc. Nên ta không cần quan tâm đến bài toán khách hàng quay trở lại mà cần quan tâm đến việc thu hút khách hàng mới.
* Bằng cách là làm tốt việc chăm sóc khách hàng khi lần đầu mua hàng và đẩy mạnh các hình thức nhằm để họ giới thiệu người thân bạn bè như tặng voucher, hoặc giảm giá nếu họ chia sẻ lên trang cá nhân, hay tag bạn bè
  1. Yếu tố khách hàng quan tâm tới sản phẩm



Thống kê mô tả các yếu tố

**Tầm quan trọng các yếu tố:**

Ta lấy độ trung bình đại diện cho tầm quan trọng của 1 yếu tố

Phân ra 3 loại: theo độ trung bình

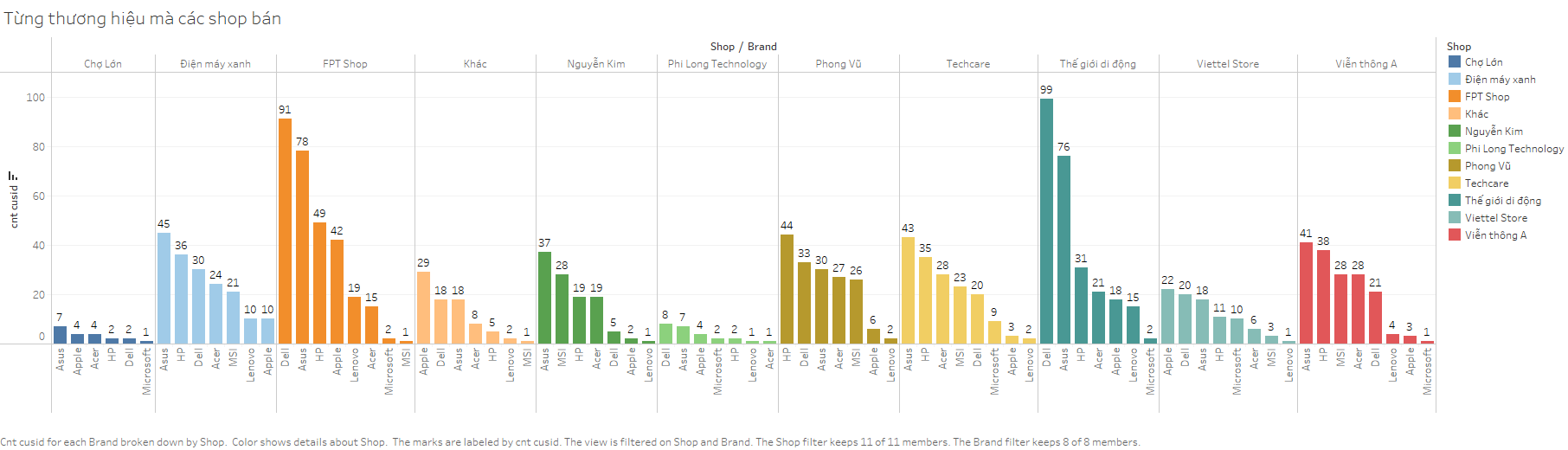
* Loại 1: từ 3.01 -3.5
* Loại 2: từ 3.5 - 3.8
* Loại 3: từ 3.8 -3.9

Yếu tố loại 1 ở mức 3. ->3.5 với mức độ quan trọng bình thường -> Gần quan trọng ta không quan tâm

Các yếu tố quan trọng: 3.8 ->3.9 với mức độ Quan Trọng là quan trọng là Yếu tố giá, Yếu tố bảo hành, Yếu tố khuyến mãi

=> Từ các yếu tố quan trọng: tập trung để nhập hàng quan tâm đến giá cũng như gia tăng chương trình bảo hành khuyến mãi để thu hút khách hàng.

## **Câu 4: Dưới góc nhìn của nhà phân tích dữ liệu, bạn có tư vấn gì cho một sàn giao dịch điện tử về mảng máy tính**



Vì đây là sàn giao dịch điện tử nên mục tiêu sẽ là bán hàng và thu về lượng khách hàng mới truy cập vào sàn => Cần phải đẩy mạnh quảng cáo để khách hàng có thể tiếp cận với sàn. FPT Shop và Thế giới di động là 2 của hàng áp dụng bán hàng trực tuyến và có chương trình khuyến mãi nhiều nhất, vì vậy đã thu về một lượng lớn khách hàng.

* 1. Một mô hình học máy đơn giản

Về việc chuẩn bị dữ liệu, lựa chọn mô hình và tính toán nhóm chúng tôi đã trình bày cụ thể trên file notebook.

Nội dung chính sẽ được trình bày cụ tỉ sau đây:

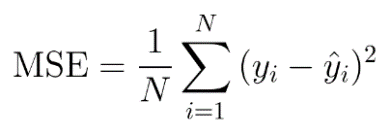
Dưới góc độ 1 nhà khoa học dữ liệu – data scientist chúng tôi sử dụng mô hình học máy đơn giản để dự đoán khoảng giá khách hàng có thể sẽ hài lòng:

Mỗi khách hàng khi tham gia chọn mua máy sẽ được nhập các dữ liệu về các yêu cầu sử dụng máy, dựa trên các dữ liệu (features) đó, từ đó mô hình học máy đưa ra một giá máy phù hợp, gợi ý cho người bán để hỗ trợ người bán tư vấn khách hàng.

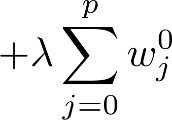
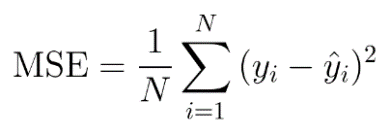
Với các features trên, một mô hình học máy đơn giản – Lasso Regression được cài đặt và chạy trên bộ dữ liệu đã có.

Dữ liệu được sử dụng như sau:

* Các features được học bao gồm: 'GioiTinh', 'Hoc\_tap', 'Lam\_viec', 'Choi\_game', 'NamHocThu', 'ChiTieuHangThang', 'ThoiGianSuDung'
* Feature được dự đoán là ‘SalePrice’
* Bộ dữ liệu được chia train-test split theo tỉ lệ 8/2.

Model dự đoán được huấn luyện trên bộ dữ liệu và đánh giá bằng điểm MSE:

Ngoài ra trong hàm mất mát (loss function) này, một regularlization(Hiệu chỉnh) term được sử dụng, với mục tiêu đơn giản hóa model, loại bỏ các feature không cần thiết trong khi học, bằng cách đưa linear-coefficient của chúng về 0:



Kết quả tốt nhất của model sau nhiều lần chạy thử nghiệm và chỉnh sửa (fine tune): **0.0364**

* 1. Đánh giá mô hình

Với dữ liệu SalePrice (sau khi đã Min-Max Scaling về khoảng giá trị 0-1) có trung bình ở 0.1936 và phương sai là 0.1575. Nhóm dự thi đánh giá kết quả của model chưa đủ tốt để sử dụng trong thực tế.

* 1. Cải tiến mô hình

Để nâng cao hiệu quả mô hình chúng tôi đã sử dụng phương pháp phân vùng dữ liệu dự đoán giá thành các khoảng giá phù hợp và sử dụng mô hình cây quyết định để đưa ra dự đoán cho các khoảng giá này

B1: Chuẩn bị dữ liệu cho mô hình

Dữ liệu được lấy như cũ ở phần mô hình trên, nhưng có sửa đổi cột SalePrice thành các nhãn 0,1,2,3,4,5,6 như sau:

Ban đầu, chúng tôi tham khảo trên trang web: https://techzones.vn/cac-laptop-van-phong-tot-nhat-trong-tung-phan-khuc-gia

|  |
| --- |
| Phân khúc tầm trung dưới 10tr |
| Phân khúc tầm cao dưới 20 tr |
| Phân khúc cao cấp 50 tr |

Phù hợp hơn chúng tôi chuyển chúng về dạng như sau: nhờ phân tích có giải thích trên file notebook

|  |  |
| --- | --- |
| Phân khúc tầm thấp dưới 5tr | 0 |
| Phân khúc tầm trung dưới 10tr | 1 |
| Phân khúc tầm cao dưới 15 tr | 2 |
| Phân khúc tầm cao dưới 20 tr | 3 |
| Phân khúc tầm cao dưới 50 tr | 4 |
| Phân khúc cao cấp dưới 50tr | 5 |
| Phân khúc cao cấp trên 50tr | 6 |

Các tham số để sử dụng huấn luyện đánh giá mô hình như cách huấn luyện thuật toán ở trên

Feature: như thuật toán được đề cập ở trên.

Target: là sale price mới được đánh category và mã hoá 0->6 như bảng trên.

B2: Đem dữ liệu huấn luyện và đánh giá, ở đây chúng tôi sử dụng tham số lựa chọn thuộc tính entropy để huấn luyện cây quyết định.

B3: Chúng tôi thu được kết quả với độ chính xác 46%

Nhóm dự thi đánh giá kết quả của model đủ tốt để sử dụng trong thực tế.

# **KẾT LUẬN**

Trên đây nhóm dự thi chúng tôi có trình bày toàn bộ quá trình phân tích dự đoán một cách chuyên sâu, sau khi dành rất nhiều thời gian xử lý dữ liệu. Thu được những tri thức quý giá và giải thích giải quyết được các vấn đề đặt ra một cách có hệ thống và đầu tư công sức, cuối cùng có trình bày 2 mô hình nhằm đưa ra dự báo dự đoán giá cho người mua giúp nhà bán hàng trên sàn thương mại gợi ý sản phẩm khi có một số thông tin từ khách hàng, chúng tôi hy vọng các tri thức và dự đoán chúng tôi nêu ra có thể giải quyết cũng như đáp ứng việc tăng doanh thu, đưa ra góc nhìn để nhà quản lý sàn thương mại có thể áp dụng những thông tin trình bày trong báo cáo vào việc ra quyết định hiệu quả, trong quá trình chúng tôi làm vẫn còn nhiều điểm hạn chế tương lai chúng tôi sẽ cập nhật thêm đào sâu vào phân tích chuyên sâu và cải thiện gia tăng hiệu quả mô hình hơn nữa.