

TEAM 4

ĐỀ BÀI: FORECASTING BUSINESS PERFORMANCE

THÀNH VIÊN:
Phạm Hồng Phúc
Đặng Tiến Dũng
Dương Tiến Thành
Phạm Tuấn Nghĩa

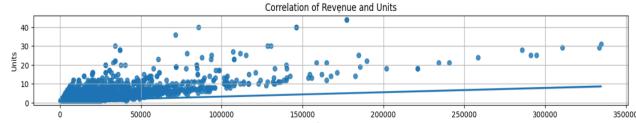
I. Phân tích khám phá dữ liệu (EDA)

1. Phân tích thống kê

a. Phân tích tương quan Revenue và Units

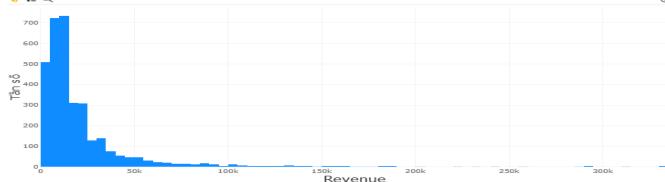
Hệ số tương quan Pearson giữa Revenue và Units là **0.325** cho thấy có mối tương quan **dương** với nhau và ở **mức độ vừa phải**. Điều này có thể thấy rõ hơn ở hình 1, với đường thẳng màu xanh là đường hồi quy tuyến tính.

Hình 1.1: Biểu đồ tương quan giữa Revenue và Units



b. Phân phối của Revenue

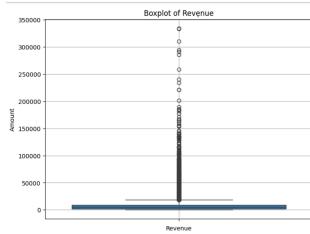
Hình 1.2: Phân phối của Revenue
Histogram của Revenue



Doanh thu có phân bố lệch phải rất nhiều -> Phần lớn giá trị doanh thu tập trung ở mức thấp, có rất ít trường hợp có doanh thu cao (phía bên phải biểu đồ có giá trị nhỏ dần). Điều này xảy ra vì có nhiều quan sát có doanh thu cao đột biến (Giá trị doanh thu ngoại lai).

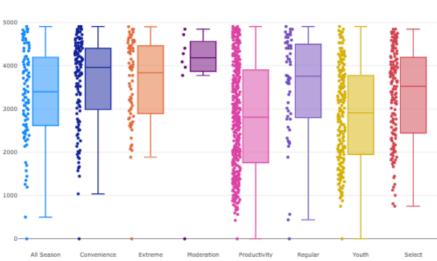
Các giá trị ngoại lai này có thể được quan sát rõ hơn ở biểu đồ Boxplot bên dưới :

Hình 1.3: Biểu đồ boxplot của Revenue



c. Phân tích Revenue theo Segment

Hình 1.4: Biểu đồ boxplot của Revenue theo Segment



Từ biểu đồ trên, ta thấy phân phối của Revenue giữa các Segment là khác nhau. Các Segment như Productivity, Youth, All Season có sự phân tán lớn, biến động cao, trong khi đó Moderation có sự dao động khá nhỏ. Tuy nhiên nhóm Moderation lại là nhóm có xu hướng doanh thu cao nhất trong các Segment.

2. Phân tích Revenue và Units theo thời gian

a. Năm

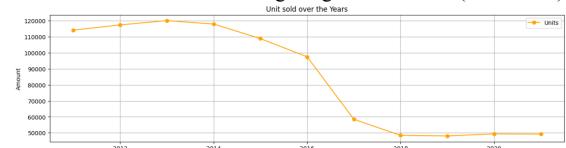
Xu hướng: Nhìn vào hình 1.5 và 1.6, ta thấy doanh thu và sản lượng bán có xu hướng **tăng** từ năm **2011 đến 2014**, sau đó **giảm mạnh**. Nguyên nhân của sự sụt giảm này có thể đến từ tình hình kinh doanh của bản thân công ty, bởi lẽ lượng hàng hóa, dịch vụ cho các mặt hàng thời trang, giày dép của Mỹ vẫn tăng trưởng ổn định (Phụ lục 1).

Từ năm **2018 đến 2021**, doanh thu và sản lượng bán **phục hồi nhẹ**. Tuy nhiên, lợi nhuận thì giảm nhẹ.

Hình 1.5 : Biểu đồ doanh thu, chi phí và lợi nhuận theo năm (2011 - 2021)



Hình 1.6: Biểu đồ số lượng hàng bán theo năm (2011-2021)



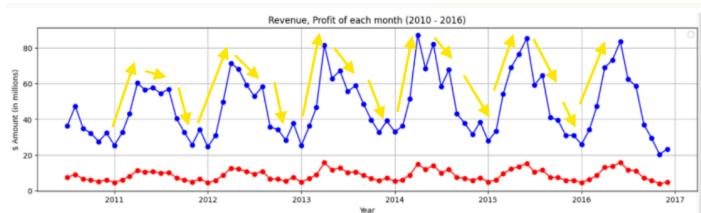
b. Tháng:

Tính mùa vụ: Dựa vào hình 1.7 và 1.8, ta thấy cả doanh thu, lợi nhuận và sản lượng bán ra đều **tăng dần** từ những **tháng đầu năm** (từ tháng 1 và tháng 6) sau đó có xu hướng **giảm dần về cuối năm** (tháng 6 đến tháng 11).

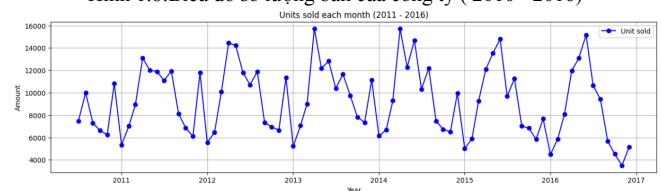
- Trong giai đoạn giảm này, có, **sự phục hồi nhẹ** ở giai đoạn **tháng 7 đến tháng 8**. Điều này xảy ra là do đây là thời điểm quay trở lại trường học của học sinh, sinh viên, họ thường mua sắm nhiều hơn để chuẩn bị cho năm học mới.
- Đặc biệt, ta thấy **có sự phục hồi mạnh vào tháng 12**, với đặc điểm là **số lượng hàng bán (Units)** **tăng mạnh**, tuy nhiên **doanh thu (Revenue)** chỉ **tăng nhẹ**, không tương xứng với mức tăng của số lượng bán. Lý do của sự phục hồi và sự không tương xứng này là do những thời điểm dịp lễ lớn như lễ Giáng Sinh, thời điểm cận Tết, người tiêu dùng có xu hướng mua sắm nhiều hơn. Và công ty cũng giảm giá hàng hóa ra để kích cầu tiêu dùng.
- Tuy nhiên, **sang tháng 1** năm tiếp theo, những yếu tố này lại **sụt giảm** do sau Tết người tiêu dùng có xu hướng thắt chặt chi tiêu.

-> **So sánh:** Quý 2 luôn là quý có các con số cao nhất trong năm, trong khi đó, quý 4 thường là quý thấp nhất so với các quý khác.

Hình 1.7: Biểu đồ doanh thu và lợi nhuận theo tháng của công ty (2010 - 2016)



Hình 1.8: Biểu đồ số lượng bán của công ty (2010 - 2016)

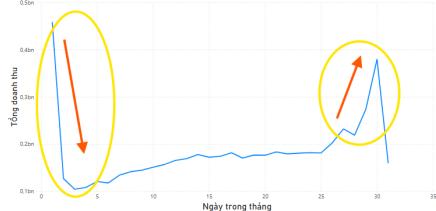


c. Ngày

Xu hướng mùa vụ: **Những ngày cuối tháng (25-30), doanh thu và sản lượng bán có xu hướng tăng mạnh**, do đây là thời điểm người tiêu dùng nhận lương, thưởng, trợ cấp (hình 1.9.)

Ngoài ra, **những ngày đầu của các tháng từ tháng 4 đến tháng 8, những con số này tăng mạnh**, có thể là do vào những dịp lễ: 01/04 - Cá tháng tư, 01/05 - Ngày Quốc tế Lao động, 01/06 - Ngày Quốc tế Thiếu nhi... người tiêu dùng có xu hướng mua sắm nhiều hơn. (Phụ lục 2)

Hình 1.9: Biểu đồ tổng doanh thu của các ngày trong tháng
Tổng doanh thu của các ngày trong tháng



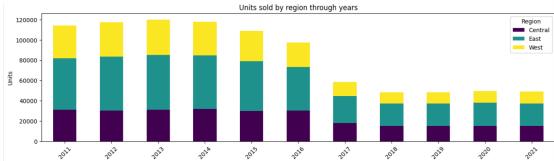
3. Phân tích revenue, units theo cơ cấu vị trí địa lý

a. Miền (Region)

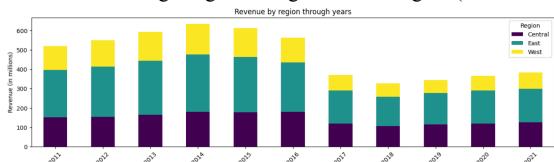
Nhìn vào hình 1.10 và 1.11, ta thấy khi xét về cơ cấu tỷ trọng, khu vực East luôn chiếm tỷ trọng lớn với số lượng bán và doanh thu dao động trong khoảng ~45-47% hàng năm.

Quan sát sự thay đổi qua thời gian, tỷ trọng ở khu vực West có xu hướng giảm nhẹ qua các năm, ngược lại, tỷ trọng khu vực Central có xu hướng tăng nhẹ.

Hình 1.10: Doanh thu hàng năm theo Region (2011-2021)



Hình 1.11: Sản lượng hàng bán hàng năm theo Region (2011 – 2021)

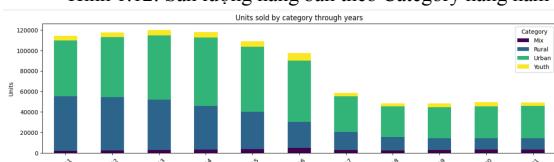


b. Đặc điểm đô thị (Category)

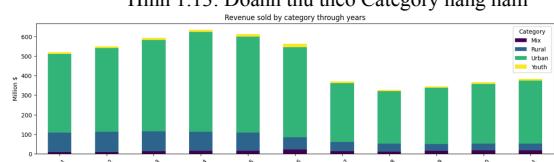
Về cơ cấu hàng bán: Sản lượng bán và doanh thu tập trung ở 2 khu vực Rural và Urban (Hình 1.12 và 1.13). Năm 2011, hai vùng này chiếm tỷ trọng hàng bán tương đương nhau, ~ 47%. Tuy nhiên qua các năm, tỷ lệ của Urban ngày càng tăng, chiếm 64.5% năm 2021, trong khi tỷ lệ của Rural năm 2021 chỉ còn 22%

Về cơ cấu doanh thu: Tuy tỷ trọng hàng bán năm 2011 của 2 khu vực Rural và Urban tương đương nhau, tuy nhiên, doanh thu ở khu vực Urban đạt 75%, trong khi khu vực Rural chỉ đạt ~ 22%, các năm sau, tỷ trọng doanh thu của 2 khu vực này có sự biến động cùng chiều với tỷ trọng hàng bán. Năm 2021, Urban chiếm 85%, trong khi Rural chỉ chiếm 8%.

Hình 1.12: Sản lượng hàng bán theo Category hàng năm



Hình 1.13: Doanh thu theo Category hàng năm



c. Phân loại theo Bang (State)

Công ty có sản phẩm bán hàng gần như toàn bộ các bang ở Mỹ, trừ bang Maine (ME) và đảo Hawaii (HI)

Các bang diện tích lớn, đông dân cư và GDP thực tế cao như CA - California (chiếm hơn 50% doanh thu và sản lượng của khu vực West (Phụ lục 3)), TX - Texas hay FL- Florida là những bang có doanh thu, số lượng hàng bán cao nhất.

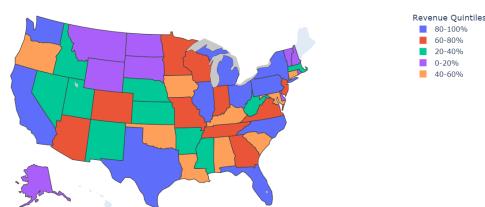
Những bang thưa dân, GDP thấp như VT - Vermont hay NH - New Hampshire là những bang có doanh thu và sản lượng bán ra thấp nhất. Ngoài ra, có một số bang có sản lượng phân vào nhóm khác nhóm doanh thu như MT - Montana hay NE - Nebraska,...

Hình 1.14: Sản lượng bán ra trung bình năm từng bang



Hình 1.15: Doanh thu trung bình năm từng bang

Average revenue by State in the US (Quintiles)



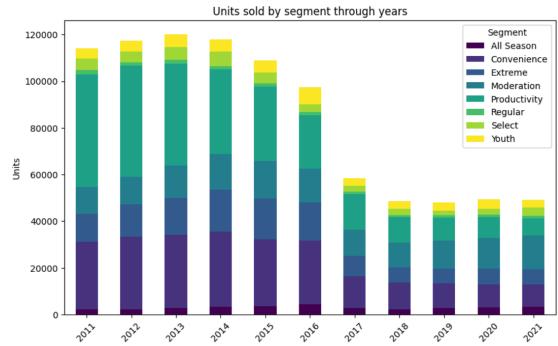
4. Phân tích theo sản phẩm

a. Kiểu sản phẩm (Segment)

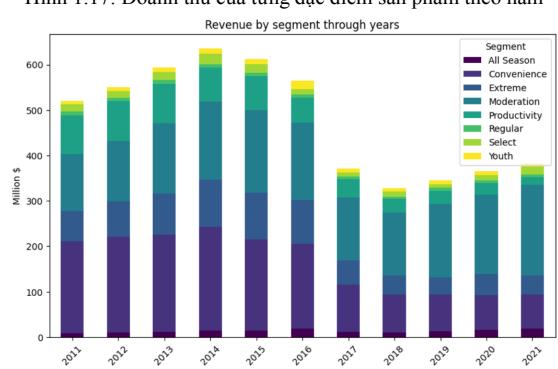
Về cơ cấu hàng bán: Bốn mặt hàng Productivity, Moderation, Extreme và Convenience chiếm phần lớn doanh thu và số lượng hàng bán, khoảng 92% (Hình 1.16 và 1.17). Những năm 2011, 2012, Productivity có số lượng bán ra nhiều nhất, khoảng 41% tổng số hàng bán. Tuy nhiên, giảm dần qua hàng năm cùng với Convenience, 2 mặt hàng còn lại có xu hướng tăng dần. Tuy nhiên, đến năm 2017, mặt hàng Extreme có dấu hiệu chững lại và giảm nhẹ. Đến năm 2019, mặt hàng Moderation chiếm tỷ trọng bán nhiều nhất, tiếp tục tăng đến nay..

Về cơ cấu doanh thu: Mặc dù Productivity những năm 2011 - 2012 có số lượng bán ra nhiều nhất, Convenience mới mang lại doanh thu nhiều nhất và cùng giảm dần theo thời gian. Cùng với xu hướng của cơ cấu hàng bán, năm 2018, mặt hàng Extreme có tỷ trọng doanh thu bắt đầu giảm nhẹ, mặt hàng Moderation chiếm tỷ trọng doanh thu cao nhất.

Hình 1.16: Sản lượng của từng đặc điểm sản phẩm theo năm



Hình 1.17: Doanh thu của từng đặc điểm sản phẩm theo năm



b. Theo dòng sản phẩm (Product Lines)

Các dòng Pirum, Natura, Maximus, Aliquis là 4 mặt hàng chiếm tỷ trọng lớn nhất trong cả doanh thu và sản lượng bán, chiếm 92 - 93% của toàn bộ công ty. Trong 12 năm, công ty có sự thay đổi, điều chỉnh về số dòng sản phẩm, ngoài ra cũng nghiên cứu thêm các dòng sản phẩm mới.

- Dòng Palma với tỷ trọng trong doanh thu, sản lượng bán ra chiếm thấp ~1%, chỉ xuất hiện năm 2010, 2011, các năm sau không còn bán nữa. Năm 2021, sản phẩm này trở lại nhưng cả doanh thu và sản lượng bán cũng chỉ ~1%.

- Dòng Salvus mới xuất hiện năm 2017, tuy nhiên doanh thu và sản lượng bán tỷ trọng không ấn tượng.

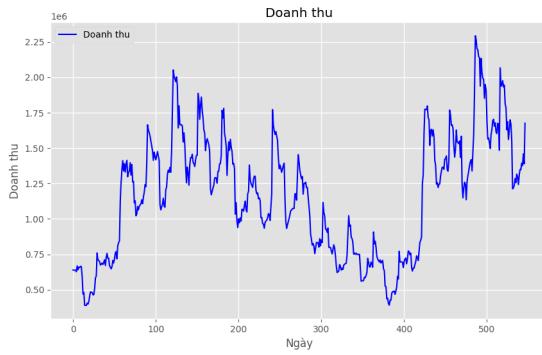
- Dòng Barba xuất hiện năm 2016, và có tỷ trọng doanh thu và sản lượng tăng nhẹ qua từng năm, tuy nhiên cũng rất bé: 1-3%

Về cơ cấu hàng bán: Natura những năm đầu có tỷ trọng lớn nhất, tuy nhiên giảm dần theo từng năm, ngược lại thì Maximus những năm đầu có tỷ trọng đứng thứ hai, tăng dần theo từng năm. Năm 2017, Maximus chiếm 28%, đứng đầu về tỷ trọng, đưa Natura về vị trí thứ hai ~21%, và khoảng cách này tăng dần sau cách năm

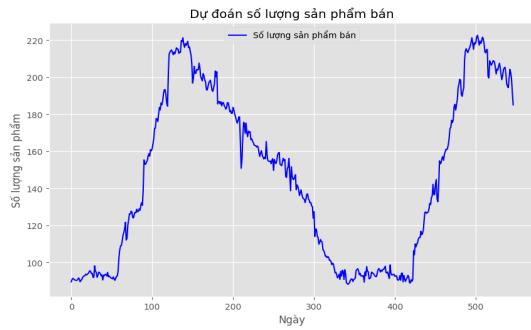
Lưu ý :

- 2 mô hình kinh tế lượng a và b được xử lý trên dữ liệu tháng, do dữ liệu ngày rất nhiều, có sự biến động cao và không thể hiện rõ được tính mùa vụ.
- Các mô hình c, d,e,f sử dụng với chuỗi dữ liệu ngày, vì các mô hình này hoạt động tốt khi cỡ mẫu lớn và có khả năng xử lý nhiễu tốt

Hình 2.1: Kết quả dự đoán Doanh thu của mô hình huấn luyện MA15 + XGBoost



Hình 2.2: Kết quả dự đoán số lượng bán của mô hình huấn luyện MA60 + XGBoost



Nhận xét về mô hình và kết quả dự báo:

- Với các mô hình kinh tế lượng : SARIMA có độ chính xác tương đối tốt so với các mô hình khác khi sử dụng dữ liệu tháng, đặc biệt trên chuỗi Revenue, trong khi đó San mũ Holt-Winters có hiệu suất kém nhất trong tất cả các mô hình, với R^2 rất thấp (0.315 cho Revenue và -0.050 cho Units). Điều này cho thấy mô hình này không phù hợp để dự báo dữ liệu này.
- XGBoost có hiệu suất khá tốt trên dữ liệu gốc. Ở chiều ngược lại, LSTM có kết quả dự báo kém.
- Khi sử dụng phương pháp MA kết hợp với XGBoost, độ chính xác được cải thiện đáng kể.
- MA15 + XGBoost cho kết quả tốt nhất trên chuỗi Revenue, với $R^2 = 0.953$ và MAPE = 0.069, chứng tỏ rằng MA15 giúp làm mịn dữ liệu tốt nhất cho dự báo doanh thu.
- MA60 + XGBoost có hiệu suất tốt nhất trên chuỗi Units, với $R^2 = 0.952$ và MAPE = 0.049, cho thấy MA60 giúp ổn định và phản ánh xu hướng dài hạn tốt nhất.
- MA7,15,30,60 + LSTM có kết quả cũng rất tốt nhưng lại không vượt qua XGBoost + MA tốt nhất -> XGBoost hoạt động hiệu quả hơn LSTM trên tập dữ liệu này, có thể do dữ liệu không quá phức tạp và XGBoost tận dụng được các biến MA tốt hơn.

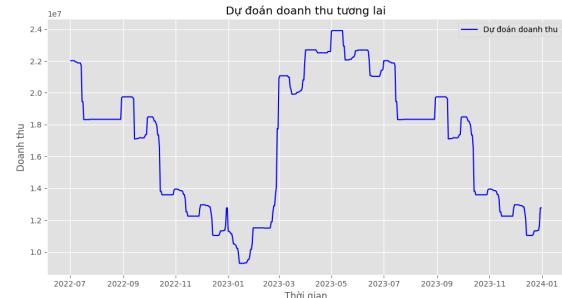
=> Tóm lại : Từ kết quả 2 bảng trên, ta thấy đối với chuỗi Revenue thì mô hình **MA15 + XGBoost** cho kết quả dự báo tốt nhất. Đối với chuỗi Units thì là **MA60 + XGBoost**. Do đó nhóm sẽ sử dụng 2 mô hình này để lập kế hoạch dự báo cho Revenue và Units trong các tháng tiếp theo.

III. Dự báo Revenue và Units từ tháng 7/2022 đến hết năm 2023

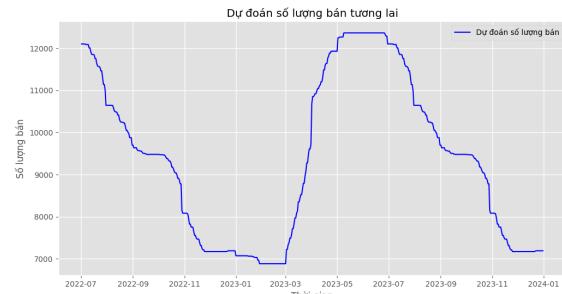
Từ hình 3.1 và 3.2, kết quả dự báo doanh thu và số lượng bán của 2 mô hình tốt nhất trong giai đoạn T7/2022 đến cuối năm 2023 cho thấy :

- Trong giai đoạn từ tháng 7 đến cuối năm, cả doanh thu và số lượng bán đều có xu hướng giảm mạnh. Điều này tương đồng về yếu tố mùa vụ đã được nói ở trên
- Giai đoạn từ đầu năm đến cuối tháng 6 thì doanh thu và số lượng bán tăng mạnh (đỉnh điểm vào tháng 5)

Hình 3.1: Kết quả dự đoán Doanh thu từ T7/2022 đến cuối năm 2023



Hình 3.2: Kết quả dự đoán số lượng bán từ T7/2022 đến cuối năm 2023



IV. Đề xuất và khuyến nghị hỗ trợ cho công ty

Dựa trên kết quả phân tích, dưới đây là các đề xuất và khuyến nghị cho công ty:

a. Yếu tố mùa vụ và xu hướng tiêu dùng:

Sự thay đổi trong doanh thu và sản lượng bán theo mùa cho thấy cần phải lập kế hoạch nhập hàng linh hoạt và chiến lược tiếp thị phù hợp với các mùa cao điểm. Cụ thể, quý 2 (từ tháng 3 đến tháng 6) có doanh thu và số lượng bán hàng cao nhất, trong khi quý 4 thường thấp nhất. Công ty cần tối ưu hóa việc nhập hàng và giảm lượng tồn kho trong những tháng thấp điểm để tránh lỗ lager.

Dịp lễ lớn, như Giáng Sinh và Tết, là thời điểm quan trọng để công ty giảm giá và kích cầu. Tận dụng những xu hướng này sẽ giúp gia tăng doanh thu.

b. Cơ cấu doanh thu và sản lượng theo địa lý và đô thị

Khu vực đô thị (Urban) đã chiếm tỷ trọng doanh thu ngày càng lớn, từ 75% năm 2011 lên 85% năm 2021. Điều này phản ánh sự thay đổi trong hành vi tiêu dùng của khách hàng ở các thành phố lớn. Công ty cần tập trung phát triển các chiến lược phân phối và marketing mạnh mẽ tại các khu vực này.

Các bang có dân cư đông đúc như California, Texas, và Florida là những khu vực có doanh thu cao nhất. Công ty nên tăng cường các chiến dịch quảng bá và phân phối sản phẩm tại những bang này để tối ưu hóa doanh thu.

Ngoài ra, để hỗ trợ tăng doanh thu, tăng lợi nhuận hiệu quả, nhóm cũng đề xuất nên tập trung vào các vùng có tỷ lệ Biên lợi nhuận (Profit Margin) cao và **tương đối ổn định (Phụ Lực 4)** đó là khu vực Youth, hoặc các bang phía Tây (West). Đây là những nơi mà có thể mang lại lợi nhuận tốt hơn các khu vực khác.

Vì East là nơi có doanh số bán hàng cao nhất, nhóm chúng tôi đề xuất nên **có kho hàng lớn ở East** để rút ngắn thời gian giao hàng.

Đây mạnh các chiến dịch quảng cáo ở các bang có doanh thu thấp, đồng thời cải thiện chiến lược phân phối tại các khu vực đang tăng trưởng mạnh như "Central" và "Urban".

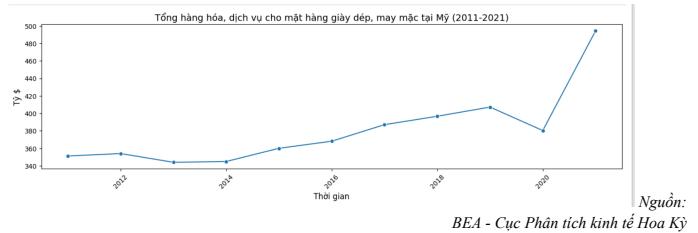
c. Cơ cấu doanh thu và sản lượng theo sản phẩm

Tăng cường sự phân bổ sản phẩm từ các dòng có xu hướng tăng trưởng, như dòng sản phẩm "Moderation", và giảm sự phụ thuộc vào những sản phẩm đang giảm dần doanh thu như "Productivity". Cập nhật và mở rộng dòng sản phẩm "Maximus" và "Moderation" để khai thác thị trường đang có nhu cầu cao hơn, đồng thời nghiên cứu thêm sản phẩm mới để duy trì sự đa dạng trong danh mục.

Ngoài ra, cần nghiên cứu thêm các xu hướng tiêu dùng mới và tăng cường các chiến dịch marketing trực tuyến để thu hút khách hàng từ các khu vực đô thị.

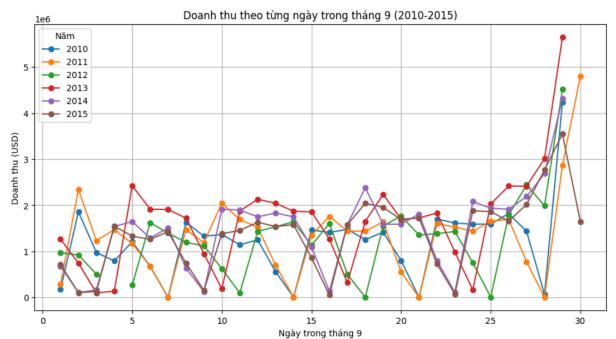
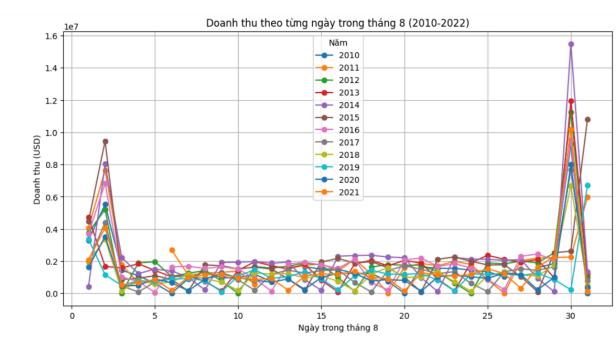
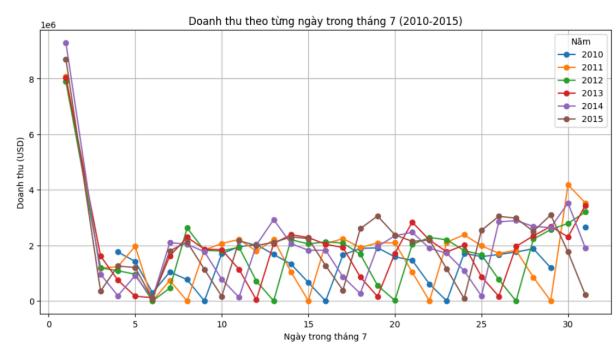
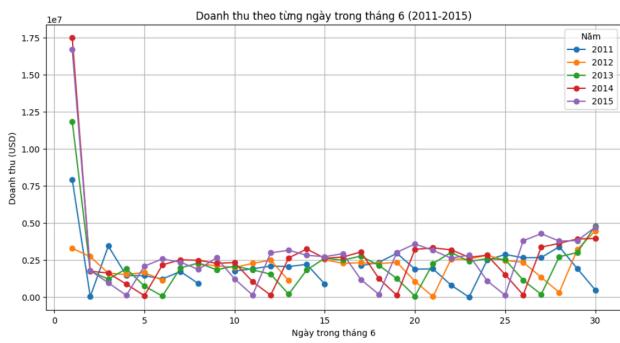
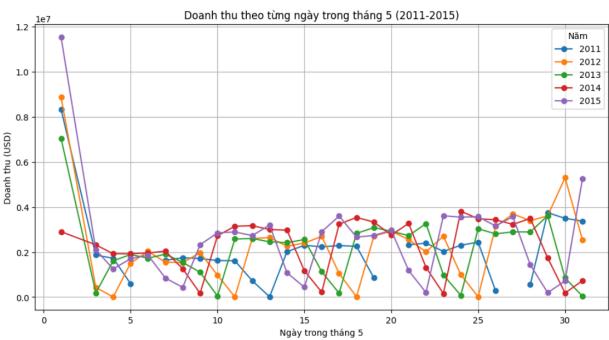
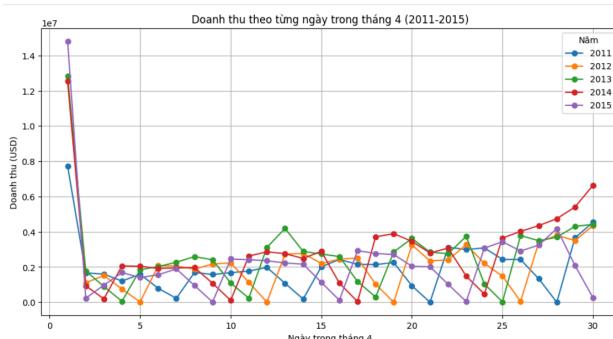
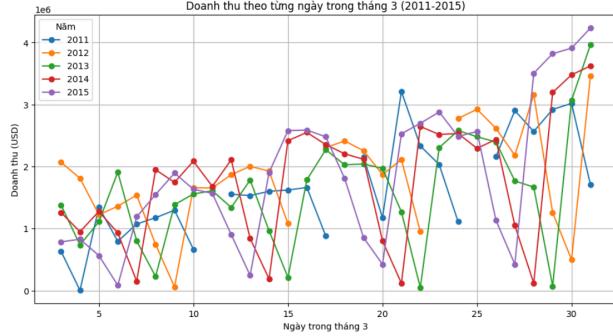
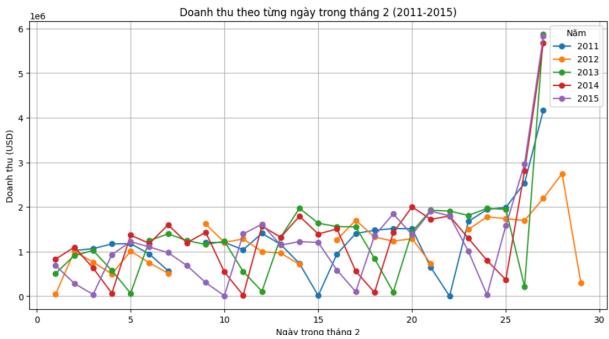
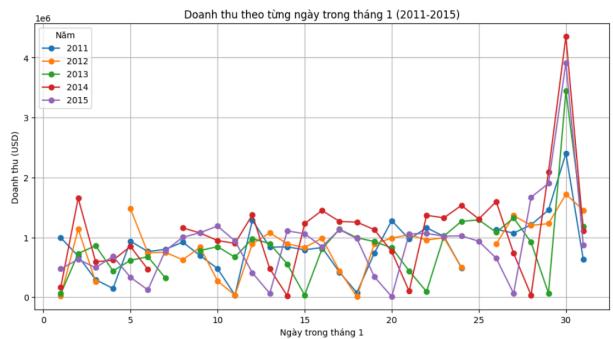
PHỤ LỤC (BIỂU ĐỒ BỔ SUNG)

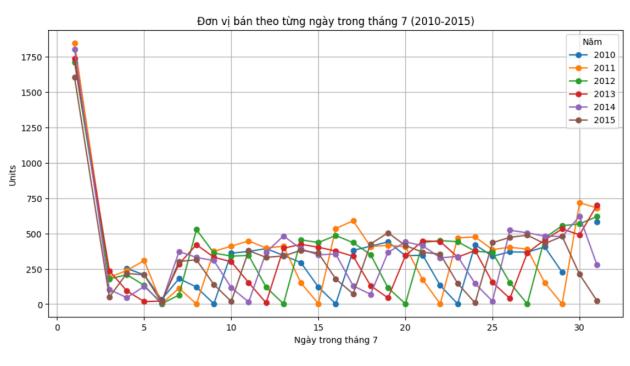
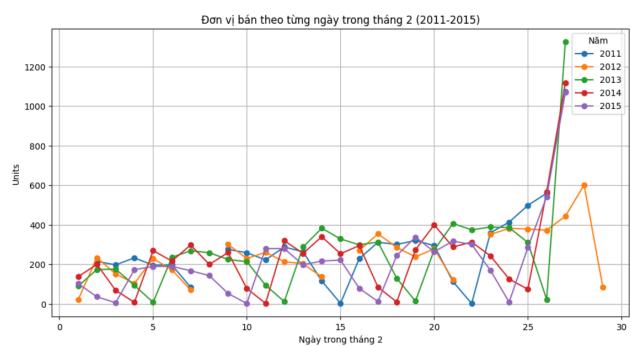
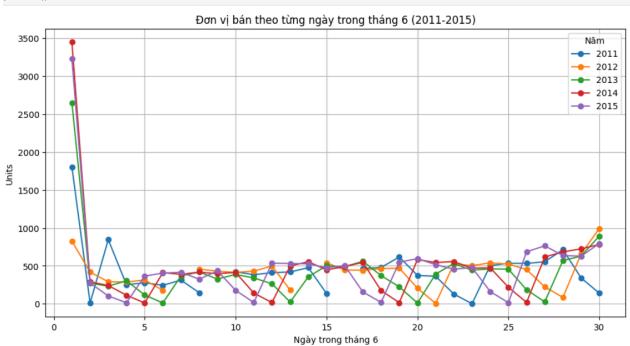
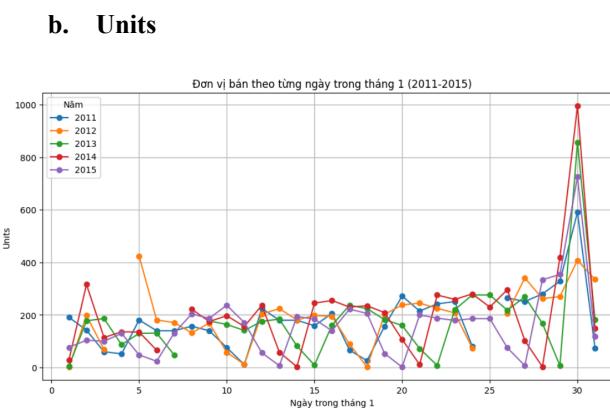
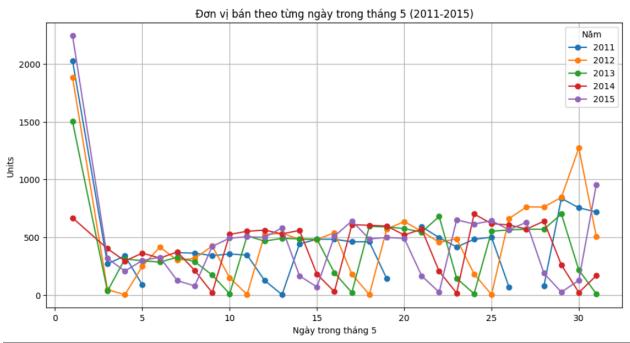
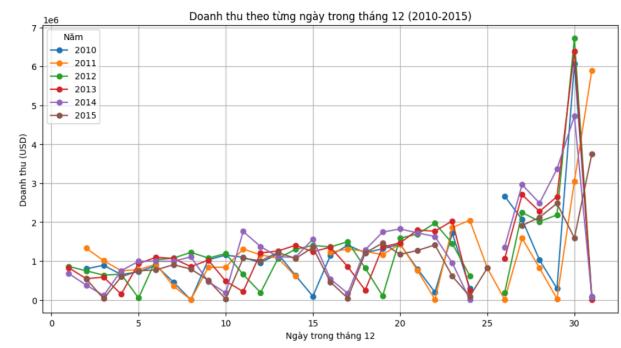
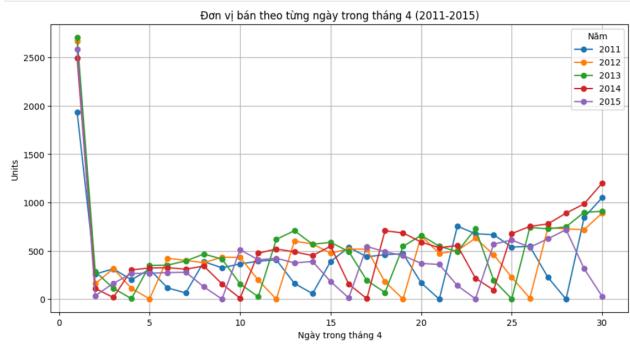
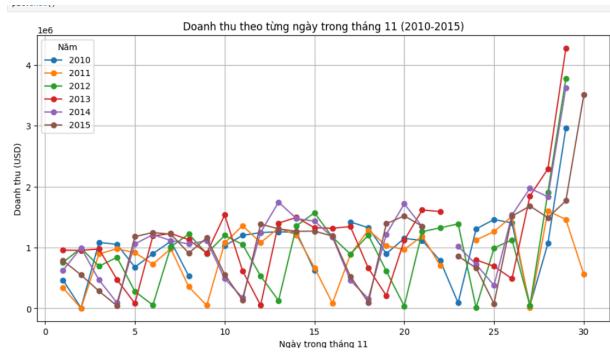
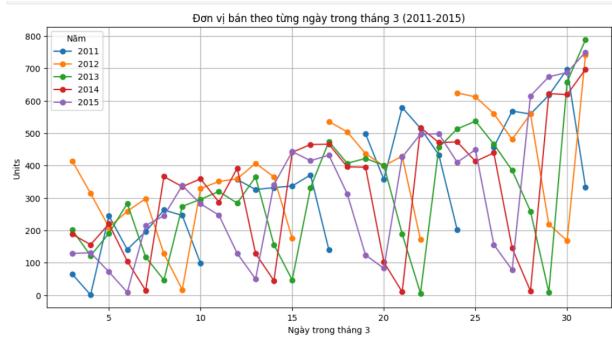
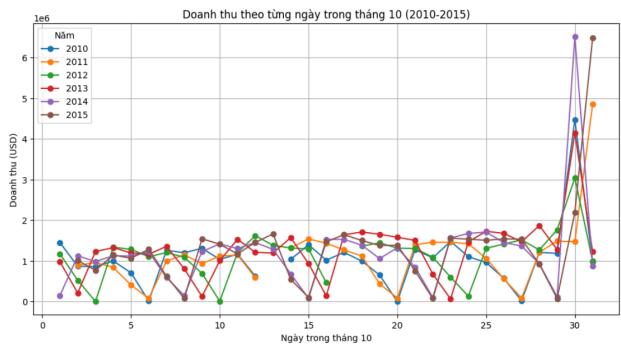
1. Hàng hóa dịch vụ cho các mặt hàng thời trang, giày dép tại Mỹ



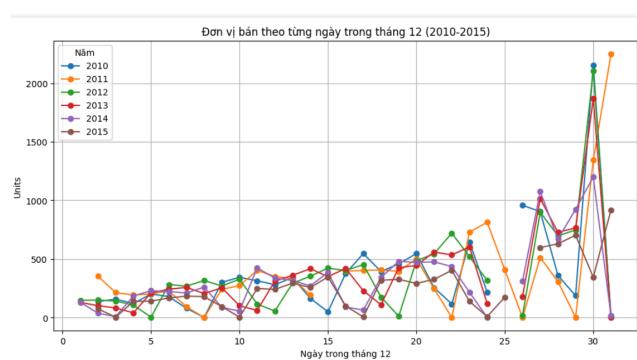
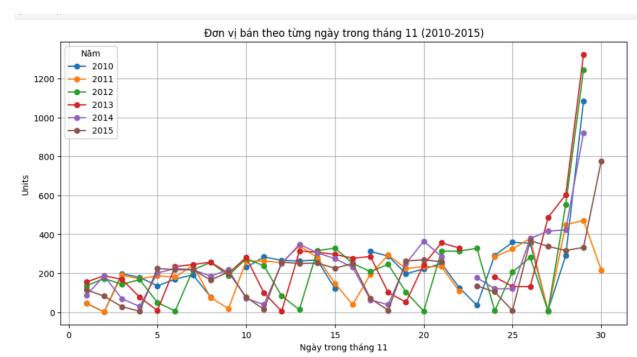
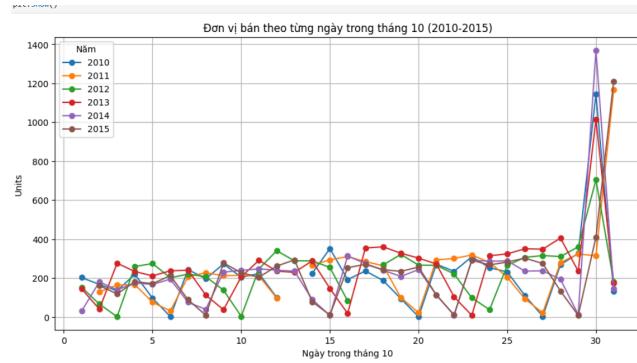
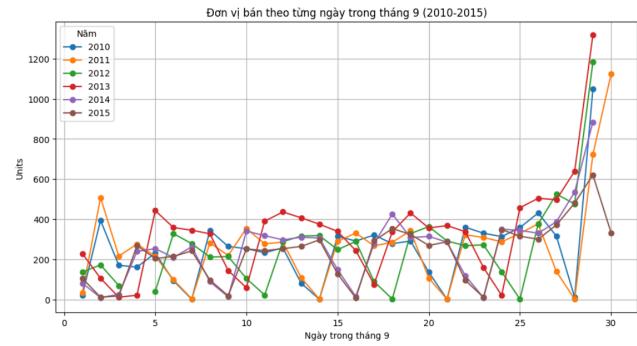
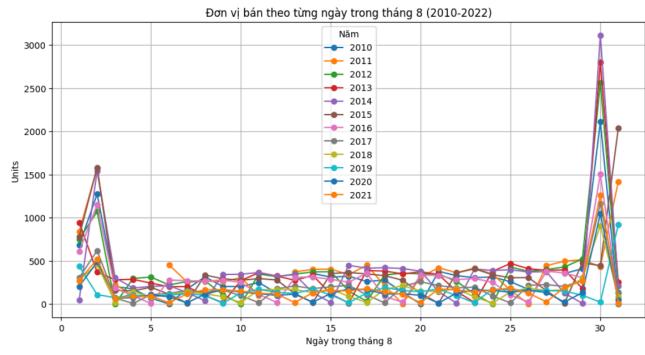
2. Doanh thu, sản lượng bán của từng ngày trong các tháng qua các năm (2011 - 2015)

a. Doanh thu (Revenue)

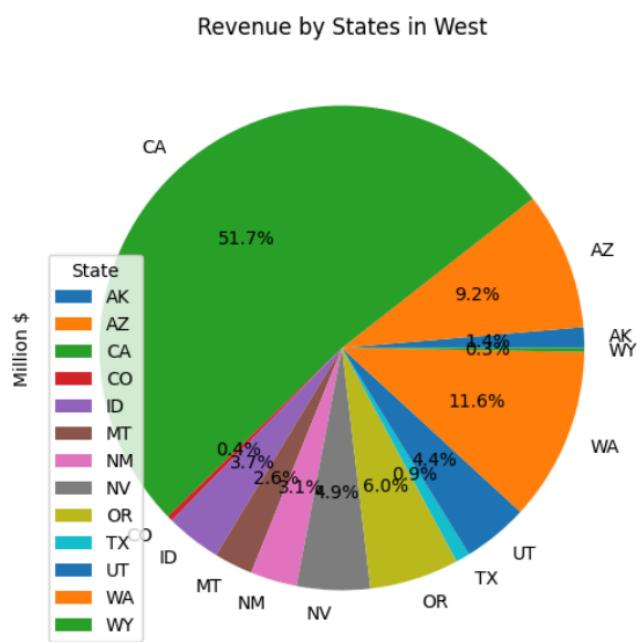
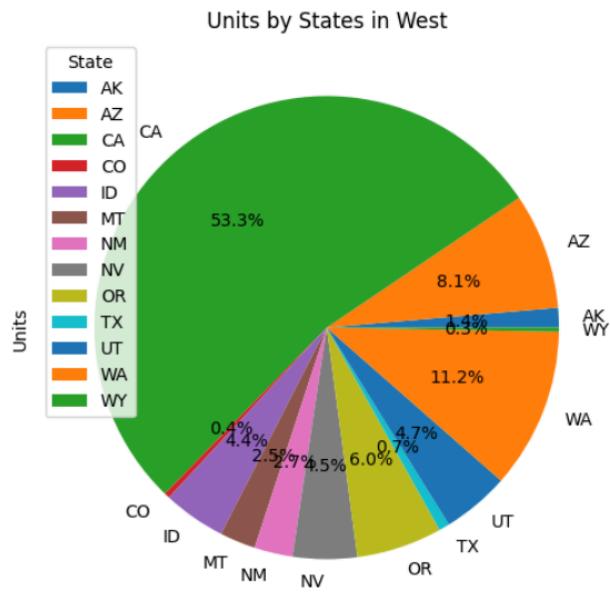




b. Units

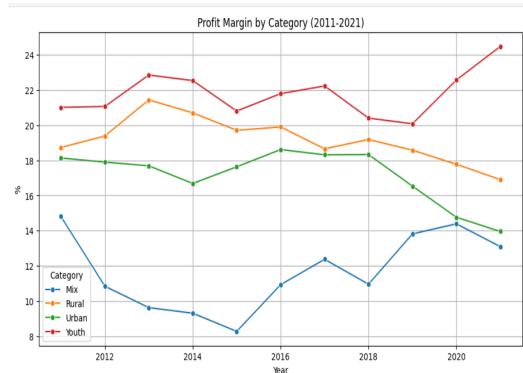


3. Tỷ trọng Doanh thu, sản lượng bán của California trong West



4. Profit margin

a. Region



b. Region

