

BÁO CÁO

Đề tài: Phân tích dữ liệu hành vi mua chéo của khách hàng

1.Mục đích:

Trong thế giới cạnh tranh khốc liệt của kinh doanh, nơi chi phí mua lại khách hàng không ngừng gia tăng, **việc giữ chân khách hàng hiện tại và tối đa hóa giá trị trọn đời** của họ là điều tối **quan trọng**. **Bán chéo** (Cross-selling) nổi lên như một chiến lược mạnh mẽ trong tình huống này, **mang lại lợi ích kép cho cả doanh nghiệp và khách hàng**.

Đối với doanh nghiệp:

- **Tăng doanh thu:** Giúp xác định các sản phẩm hoặc dịch vụ mà khách hàng có xu hướng mua cùng nhau.

→ Thiết kế các chương trình khuyến mãi hoặc gợi ý sản phẩm để tăng giá trị đơn hàng.

- **Cá nhân hóa trải nghiệm khách hàng:** Đưa ra các gợi ý phù hợp hơn với nhu cầu và sở thích của từng khách hàng, từ đó cải thiện sự hài lòng và giữ chân khách hàng.
- **Tối ưu hóa danh mục sản phẩm:** Phân tích giúp nhận ra các nhóm sản phẩm thường được mua kèm, từ đó doanh nghiệp có thể điều chỉnh chiến lược sắp xếp hàng hóa hoặc thiết kế các gói sản phẩm phù hợp.
- **Dự báo nhu cầu:** Hiểu được xu hướng mua hàng của khách hàng giúp dự đoán các sản phẩm có khả năng bán chạy, từ đó tối ưu hóa chuỗi cung ứng và quản lý tồn kho.
- **Giảm chi phí marketing:** Thay vì quảng bá mọi sản phẩm đến tất cả khách hàng, doanh nghiệp chỉ cần tập trung vào nhóm sản phẩm phù hợp với từng đối tượng, giúp tiết kiệm chi phí và tăng hiệu quả chiến dịch.
- **Xây dựng chiến lược bán hàng hiệu quả hơn:** Phân tích mua chéo cung cấp thông tin để tạo ra các chương trình ưu đãi, chính sách giá hoặc gợi ý thông minh, giúp nâng cao tỷ lệ chuyển đổi và giá trị vòng đời khách hàng.

Đối với khách hàng:

- **Tăng tiện ích và tiết kiệm thời gian:** Tìm thấy các mặt hàng liên quan mà không cần mất thời gian tìm kiếm.
- **Trải nghiệm mua sắm cá nhân hóa:** Nhận được các đề xuất được tùy chỉnh phù hợp với sở thích và thói quen tiêu dùng của họ, mang lại cảm giác được quan tâm.
- **Cơ hội nhận ưu đãi hấp dẫn:** cung cấp các gói combo hoặc khuyến mãi đặc biệt cho các sản phẩm thường mua kèm, giúp khách hàng tiết kiệm chi phí.

- **Tăng giá trị và hiệu quả sử dụng sản phẩm:** Gợi ý mua chéo thường nhắm đến các sản phẩm bổ trợ hoặc liên quan, giúp khách hàng sử dụng sản phẩm hiệu quả hơn.
- **Phát hiện nhu cầu tiềm ẩn:** Được gợi ý những sản phẩm mà họ chưa nghĩ đến nhưng thực sự hữu ích, giúp họ khám phá thêm lựa chọn phù hợp với nhu cầu.
- **Cải thiện sự hài lòng và trung thành:** Nhờ nhận được trải nghiệm tốt hơn, khách hàng cảm thấy hài lòng và có xu hướng quay lại mua sắm tại doanh nghiệp.

Nhận xét: Phân tích dữ liệu mua chéo không chỉ nhằm **tăng doanh số** mà còn **giúp doanh nghiệp hiểu rõ hơn về hành vi và nhu cầu của khách hàng**. Từ đó, **nâng cao trải nghiệm mua sắm của khách hàng** mang lại **sự tiện lợi và tiết kiệm**, đồng thời đáp ứng tốt hơn nhu cầu của họ và **xây dựng các chiến lược kinh doanh bền vững**.

2. Hướng ứng dụng:

*Có thể **áp dụng trong nhiều lĩnh vực khác nhau**, một số lĩnh vực tiêu biểu:

- Phân tích hành vi mua chéo của khách hàng có thể được ứng dụng trong nhiều lĩnh vực khác nhau, **đặc biệt là những ngành tập trung vào việc cải thiện trải nghiệm khách hàng và tối ưu hóa doanh thu**. Dưới đây là các lĩnh vực tiêu biểu:
 - **Thương mại điện tử:**
 - Gợi ý sản phẩm hoặc dịch vụ liên quan dựa trên lịch sử mua sắm của khách hàng.
 - Tạo các chương trình khuyến mãi hoặc combo sản phẩm nhằm tăng giá trị giỏ hàng.
 - **Bán lẻ**
 - Tối ưu hóa sắp xếp sản phẩm trong cửa hàng vật lý để tăng khả năng mua chéo.
 - Xây dựng các gói sản phẩm thường xuyên được mua cùng nhau.
 - **Ngân hàng và tài chính**
 - Đề xuất các sản phẩm tài chính bổ sung như bảo hiểm, thẻ tín dụng, hoặc khoản vay dựa trên nhu cầu của khách hàng.
 - Gợi ý dịch vụ nâng cấp tài khoản hoặc đầu tư cho khách hàng sử dụng dịch vụ tiết kiệm.
 - **Du lịch và khách sạn**
 - Gợi ý các dịch vụ bổ sung như vé tham quan, phương tiện đưa đón, hoặc bảo hiểm du lịch.
 - Tạo các gói dịch vụ trọn gói (combo) như vé máy bay, phòng khách sạn, và tour du lịch.
 - **Chăm sóc sức khỏe**

→ Gợi ý các sản phẩm y tế liên quan, như thuốc bổ sung, dụng cụ y tế, hoặc dịch vụ xét nghiệm sức khỏe.

→ Tăng cường dịch vụ tư vấn sức khỏe cá nhân hóa dựa trên lịch sử điều trị.

- **Công nghệ và phần mềm**

→ Đề xuất các phần mềm hoặc dịch vụ bổ sung liên quan đến sản phẩm khách hàng đã mua.

→ Tạo các gói dịch vụ tích hợp như phần mềm và hỗ trợ kỹ thuật.

- **Giáo dục**

→ Gợi ý khóa học bổ trợ hoặc tài liệu học tập liên quan đến khóa học chính.

→ Tạo các gói ưu đãi học tập trọn gói bao gồm nhiều khóa học cùng lĩnh vực.

- **Giải trí và truyền thông**

→ Gợi ý các dịch vụ hoặc sản phẩm bổ sung như vé xem phim, gói dịch vụ streaming, hoặc sản phẩm liên quan đến nội dung yêu thích của khách hàng.

- **Dịch vụ ô tô**

→ Gợi ý các dịch vụ bổ sung như bảo dưỡng định kỳ, lắp đặt phụ kiện, hoặc bảo hiểm xe.

- **Ngành thực phẩm và đồ uống**

→ Đề xuất các món ăn hoặc đồ uống thường được kết hợp cùng nhau.

→ Tạo các combo đồ ăn hoặc dịch vụ giao hàng trọn gói.

Kết luận: Đối với nguồn dữ liệu nhóm em đã chọn hiện tại, hướng lĩnh vực nhóm em muốn hướng tới và áp dụng là lĩnh vực thương mại điện tử, bán lẻ và ngành thực phẩm ăn uống

Về quy mô cửa hàng:

- Cửa hàng nhỏ: Tập trung món phổ biến, khuyến mãi nhỏ, tối ưu không gian.
- Cửa hàng trung bình: Đẩy mạnh combo nhóm nhỏ, quản lý giờ cao điểm.
- Cửa hàng lớn: Thiết kế combo lớn, tập trung vào trải nghiệm nhóm, và sự kiện.

3.Phương pháp phân tích: Phân tích dự đoán và phân tích đề xuất

1. Phân tích dự đoán (Predictive Analysis): Trả lời cho câu hỏi "Điều gì có thể xảy ra?"
 - Xu hướng kết hợp sản phẩm: Một số sản phẩm có xu hướng được mua cùng nhau thường xuyên.
 - Tần suất mua hàng lặp lại: Dự đoán rằng các món phổ biến sẽ tiếp tục có nhu cầu cao, đặc biệt nếu khách hàng thường xuyên mua sắm.
 - Thời gian và dịp mua hàng: Có thể có các khoảng thời gian cụ thể (giờ cao điểm hoặc ngày lễ) khi các món như combo hoặc món chính sẽ tăng mạnh.
2. Phân tích đề xuất (Prescriptive Analysis): Trả lời cho câu hỏi "Chúng ta nên làm gì tiếp theo?".

- Thiết kế chương trình khuyến mãi và combo, gợi ý combo theo xu hướng mua hàng, đặt tên các combo hấp dẫn
- Khuyến mãi sản phẩm kèm theo: Áp dụng giảm giá cho các món phụ khi mua món chính
- Tối ưu hóa menu: Sắp xếp menu để làm nổi bật các món phổ biến hoặc các món được kết hợp nhiều nhất.
- Quản lý trải nghiệm tại cửa hàng: Sắp xếp quầy hàng hoặc khu vực chế biến để dễ dàng đáp ứng các combo phổ biến, giảm thời gian chờ đợi của khách hàng.
- Dự trữ hàng tồn kho hợp lý đối với những combo khuyến mãi hoặc món ăn được chuộng nhất của hàng

4.Nguồn dữ liệu:

- Tập dữ liệu mà nhóm chúng em chọn có tên là "Chipotle.tsv" và được lưu trữ trên trang Kaggle. Nguồn dữ liệu thu thập:
<https://www.kaggle.com/datasets/navneethc/chipotle>
- Tập dữ liệu này chứa thông tin chi tiết về đơn hàng từ một nhà hàng bao gồm ID đơn hàng, số lượng đã đặt, tên món, lựa chọn, giá cả và tổng số. Nhiều loại món ăn chính, đồ uống và đồ ăn kèm được đặt với nhiều loại protein, nước sốt và topping kết hợp.
- Bộ dữ liệu gồm doanh số bán hàng trong ngày của Chipotle

5.Lý do chọn đề tài:

- **Tính thực tiễn cao:** Giúp doanh nghiệp tăng doanh thu, cải thiện trải nghiệm khách hàng, và tối ưu hóa chiến lược bán hàng.
- **Xu hướng thời đại:** Phù hợp với xu hướng cá nhân hóa và ứng dụng công nghệ như Big Data, AI, Machine Learning.
- **Tính đa ngành:** Ứng dụng được trong nhiều lĩnh vực
- Góp phần **nâng cao hiệu quả kinh doanh:** Hỗ trợ ra quyết định chính xác về giá, khuyến mãi, và quản lý sản phẩm/dịch vụ.
- **Cơ hội nghiên cứu sáng tạo:** Khai thác sâu hành vi khách hàng và mở rộng ứng dụng công nghệ hiện đại.

6.Đề xuất và Kết luận:

**Đề xuất:*

**** Combo dựa trên hành vi mua chéo tích cực**

Tận dụng các món có liên kết mạnh mẽ để thiết kế combo thu hút, giúp tăng khả năng chọn mua.

A. Combo phổ biến:

1. Combo "Tăng cường trải nghiệm":

○ **Chicken Bowl + Chips + Canned Soft Drink/Veggie Soft Tacos**

Giải thích:

Chicken Bowl: Là một trong các món chính có hiệu suất dự đoán cao (Accuracy 81%, Precision 82%, Recall 79%), rất phổ biến.

Chips: Có hệ số hồi quy dương cao nhất với Chicken Bowl (~0.55), cho thấy sự bổ trợ mạnh khi xuất hiện cùng nhau.

Canned Soft Drink/Veggie Soft Tacos: Hệ số hồi quy tích cực, đặc biệt là liên kết tốt với nhiều món như Chicken Bowl và Veggie Burrito, giúp tăng khả năng khách hàng chọn combo này.

→ Tối ưu hóa doanh thu từ món chính phổ biến, kết hợp món phụ có liên kết mạnh. Phù hợp với khách hàng ưa thích sự đơn giản, dễ mua.

○ **Steak Bowl + Chips + Bottled Water**

Giải thích:

Steak Bowl: Đạt độ chính xác 86% với Precision và Recall cao, cho thấy món này rất phổ biến. Biểu đồ cũng cho thấy các món liên quan như Chips có ảnh hưởng tích cực.

Chips: Là món phụ liên kết tích cực mạnh với Steak Bowl, tăng khả năng dự đoán chọn món chính.

Bottled Water: Là lựa chọn trung tính, ít ảnh hưởng tiêu cực và thích hợp để kết hợp với món chính, đặc biệt cho nhóm khách hàng muốn lựa chọn lành mạnh hơn.

2. Combo "Thỏa mãn vị giác": **Chicken Bowl + Chips and Guacamole + Canned Soda**

Giải thích:

- **Chicken Bowl:** Là một trong các món chính có hiệu suất tốt nhất với Accuracy 81% và F1-score cao. Biểu đồ chỉ ra món này có mối liên kết tích cực với Chips.
- **Chips and Guacamole:** Có liên kết tích cực rõ ràng với nhiều món chính như Chicken Bowl, phù hợp để kết hợp.

- **Canned Soda:** Kết quả mô hình cho thấy đây là món có Precision và Recall khá cao, phổ biến và dễ thêm vào thực đơn.

3. Combo "Đặc biệt lành mạnh":

Veggie Soft Tacos + Veggie Bowl + Bottled Water

Giải thích:

- Veggie Bowl: Accuracy đạt 86.47%, phù hợp với nhóm khách hàng quan tâm đến thực phẩm lành mạnh.
- Veggie Soft Tacos: Hệ số hồi quy dương cao nhất với Veggie Bowl, hỗ trợ mạnh mẽ trong việc tăng khả năng chọn Veggie Bowl.
- Bottled Water: Là lựa chọn tự nhiên cho nhóm khách hàng này, không cạnh tranh với các sản phẩm lành mạnh.

→ Tăng giá trị đơn hàng cho khách hàng chọn sản phẩm lành mạnh. Nhắm đến khách hàng ưu tiên sức khỏe, thực phẩm ít dầu mỡ.

4. Combo "Năng động":

Chicken Burrito/Steak Burrito + Canned Soda + Chips and Tomatillo Red Chili Salsa

Giải thích:

- Chicken Burrito/Steak Burrito: Độ chính xác cao, phù hợp với khách hàng thích món đậm đà, nhiều năng lượng.
 - Chips and Tomatillo Red Chili Salsa: Có hệ số hồi quy tích cực với nhiều món chính, đặc biệt hỗ trợ Steak Burrito.
 - Canned Soda: Đồ uống phổ biến, tăng khả năng khách hàng chọn combo khi tìm kiếm bữa ăn nhanh gọn.
- Phù hợp cho khách trẻ, năng động, thích đồ ăn nhanh.

5. Combo "Truyền thống hoàn hảo": Chicken Burrito + Chips + Canned Soft Drink

- **Giải thích:**
 - **Chicken Burrito:** Độ chính xác đạt 82%, hiệu suất cao, được nhiều khách hàng lựa chọn.

- **Chips** : Hệ số hồi quy tích cực với Chicken Burrito, giúp tăng khả năng lựa chọn.
- **Canned Soft Drink**: Phù hợp để bổ sung vị và tối ưu doanh thu.

→ Hút khách hàng truyền thống với lựa chọn dễ chấp nhận và quen thuộc.

**** Combo nhắm đến lợi nhuận cao:**

1. Combo "Siêu kích thích vị giác":

Steak Bowl + Chips and Guacamole + Canned Soda

→ Nhấn mạnh hương vị độc đáo của Steak Bowl, đi kèm với món phụ mạnh.

2. Combo "Truyền thống cải tiến": Chicken Salad Bowl + Chips and Tomatillo Green Chili Salsa + Canned Soft Drink

Giải thích:

- Chicken Salad Bowl: Là món chính phổ biến, đạt độ chính xác 87.3% trong mô hình với F1-score cao (~0.87 cho cả hai lớp).

- Chips and Tomatillo Green Chili Salsa: Hệ số hồi quy dương cao với các món chính như Chicken Bowl và Bottled Water, cho thấy khả năng bổ trợ tốt.

- Canned Soft Drink: Đồ uống phổ biến, dễ phối hợp, phù hợp với cả món chính nhẹ như Chicken Salad Bowl và salsa cay.

→ Kết hợp món truyền thống với sự phá cách từ salsa cay.

- **Crispy Tacos + Chips + Bottled Water**

→ Tạo trải nghiệm đơn giản, tiết kiệm nhưng vẫn đảm bảo tính hấp dẫn.

**** Combo giảm thiểu món có tác động tiêu cực**

Nhằm hạn chế tác động của các món làm giảm xác suất chọn món chính.

1. Combo điều chỉnh:

- **Canned Soft Drink + Chicken Bowl + Veggie Crispy Tacos**

Chicken Bowl: Phổ biến, có độ chính xác cao (81%).

Canned Soft Drink: Có hệ số hồi quy dương, bổ trợ Chicken Bowl tốt.

Veggie Crispy Tacos: Thay vì để riêng lẻ (tác động tiêu cực trong một số trường hợp), đưa vào combo để giảm ảnh hưởng xấu.

→ Thay vì quảng bá riêng Veggie Crispy Tacos, đưa vào combo để giảm ảnh hưởng tiêu cực của món này.

- **Steak Salad + Veggie Burrito + Chips**

Veggie Burrito: Độ chính xác cao nhất (93%), là món chính ưu tiên.

Steak Salad: Tác động tiêu cực đến nhiều món chính, được đưa vào làm món phụ để giảm cạnh tranh.

Chips: Tăng giá trị combo, hệ số hồi quy tích cực

→ Định vị Steak Salad làm món phụ, không cạnh tranh trực tiếp với các món chính.

2. Combo thay thế:

- **Veggie Soft Tacos + Canned Soda + Chips**

Veggie Soft Tacos: Tích cực nhất với các món chính (hệ số hồi quy cao).

Canned Soda + Chips: Bổ trợ phổ biến, phù hợp trong mọi combo.

→ Sử dụng món phụ Veggie Soft Tacos để thay thế Crispy Tacos trong thực đơn.

- **Chicken Salad Bowl + Chips and Tomatillo Green Chili Salsa**

Chips and Tomatillo Green Chili Salsa: Hệ số hồi quy dương, giảm cạnh tranh trực tiếp từ Salad truyền thống.

Chicken Salad Bowl: Độ chính xác cao (87%), đảm bảo sức hút chính.

→ Thay thế Salad truyền thống bằng món phụ Chips and Tomatillo để giảm sự cạnh tranh.

3. Gợi ý chiến lược quảng bá

1. Combo theo khung giờ:

- Bữa trưa: **Chicken Burrito + Chips + Bottled Water/ Canned Soft Drink**
→ Giá ưu đãi hơn 10% trong giờ trưa.
- Bữa tối: **Steak Bowl + Chips and Guacamole + Canned Soft Drink**
→ Tặng tính hấp dẫn cho khách hàng ăn tối.

2. Combo theo nhóm:

- **Combo gia đình (4 người):**
 - 2 Chicken Bowls + 2 Veggie Burritos + Chips and Guacamole + Bottled Water (4 suất).
→ Giảm giá 15% khi mua cả set.
- **Combo bạn bè (2 người):**
 - 1 Steak Burrito + 1 Veggie Bowl + 2 Canned Sodas + Chips.
→ Tặng kèm Chips and Salsa khi mua combo này.

3. Combo theo mùa:

- **Hè mát lạnh:** Canned Soft Drink + Chips and Mild Fresh Tomato Salsa + Chicken Salad Bowl.
→ Quảng bá trong mùa nóng, nhấn mạnh tính giải nhiệt.

**Kết luận:*

Hiệu suất mô hình: Logistic regression hoạt động hiệu quả trong phân tích hành vi mua chéo, với độ chính xác từ 77% đến 93%. Các kết quả chỉ ra mối quan hệ mạnh mẽ giữa các sản phẩm, giúp xác định các combo tối ưu.

Chiến lược thiết kế combo:

- Tập trung vào các combo có liên kết mạnh mẽ như **Chicken Bowl + Chips, Veggie Soft Tacos + Veggie Bowl** để tối ưu doanh thu.
- Tập trung quảng bá combo có Chips và Canned Soft Drink (hoặc Bottled Water).
- Tận dụng combo giảm thiểu ảnh hưởng của các sản phẩm có tác động tiêu cực như Crispy Tacos hay Steak Salad.

Tối ưu doanh số:

- Quảng bá combo theo **khung giờ, nhóm đối tượng, hoặc mùa vụ** để đáp ứng nhu cầu cụ thể.
- Kết hợp các sản phẩm có tính bổ trợ, tăng tỷ lệ gắn bó và giá trị hóa đơn trung bình.

Chiến lược phát triển thực đơn:

- Đa dạng hóa món ăn để giảm cạnh tranh nội bộ (ví dụ: thêm các thành phần đặc biệt cho Chicken Salad Bowl).
- Cải thiện món có tác động tiêu cực để gia tăng hiệu quả mua hàng chéo.

Ứng dụng thực tế: Các đề xuất có thể áp dụng trực tiếp để tăng doanh thu và cải thiện trải nghiệm khách hàng thông qua các combo tối ưu và chiến lược khuyến mãi phù hợp.

7. Mô tả quá trình phân tích

I. Khám phá dữ liệu

- **Mục tiêu:** Hiểu rõ cấu trúc và nội dung của dữ liệu

1. Dữ liệu gồm 4622 dòng và 5 cột:

- order_id: Mã đơn hàng.
- quantity: Số lượng của mặt hàng trong đơn hàng.
- item_name: Tên mặt hàng.
- choice_description: Mô tả lựa chọn tùy chỉnh (nếu có).
- item_price: Giá của mặt hàng (dạng chuỗi, bao gồm ký hiệu \$).

```
Data columns (total 5 columns):
#   Column                Non-Null Count  Dtype
---  -
0   order_id               4622 non-null   int64
1   quantity               4622 non-null   int64
2   item_name              4622 non-null   object
3   choice_description     3376 non-null   object
4   item_price             4622 non-null   object
```

2. Bài toán cần phân tích:

- Các sản phẩm phổ biến nhất, thường xuyên được mua chung với các sản phẩm khác.
- Các sản phẩm phụ được kết hợp nhiều nhất, thường đi kèm với món chính để tăng trải nghiệm.
- Các đơn hàng có xu hướng mua lặp lại cùng một sản phẩm với số lượng lớn.

3. Các biến quan tâm:

- order_id: Dùng để nhóm các sản phẩm mua chung trong một giao dịch. Đây là biến cơ sở để xác định mối quan hệ giữa các sản phẩm.
- item_name: Tên sản phẩm để nhận diện các món hàng cụ thể trong phân tích.
- choice_description: Hiểu rõ các biến thể hoặc tùy chọn mà khách hàng ưa thích khi mua hàng.

- quantity: Số lượng của mỗi sản phẩm trong từng đơn hàng. Dùng để xác định mức độ phổ biến của sản phẩm hoặc tính tổng số lượng trong các nhóm sản phẩm.
- item_price: Giá trị sản phẩm, giúp đánh giá sự kết hợp dựa trên giá trị kinh tế (ví dụ: sản phẩm cao cấp có xu hướng kết hợp với sản phẩm nào?).

Ngoài ra, ta cần tạo thêm 1 số biến phục vụ cho việc phân tích:

- Danh sách sản phẩm trong từng đơn hàng: Dùng để xác định các cặp sản phẩm hoặc nhóm sản phẩm xuất hiện cùng nhau.
- Tần suất xuất hiện (frequency): Số lần mỗi sản phẩm hoặc mỗi cặp sản phẩm xuất hiện.
- Doanh thu theo cặp sản phẩm: Tổng giá trị của các sản phẩm khi mua chung để tìm ra các cặp tạo ra giá trị cao.

II. Tiền xử lý dữ liệu

- **Mục tiêu:** Làm sạch dữ liệu để phục vụ cho việc phân tích.

```
df.describe(include='all')
```

	order_id	quantity	item_name	choice_description	item_price
count	4622.000000	4622.000000	4622	3376	4622
unique	NaN	NaN	50	1043	78
top	NaN	NaN	Chicken Bowl	[Diet Coke]	\$8.75
freq	NaN	NaN	726	134	730
mean	927.254868	1.075725	NaN	NaN	NaN
std	528.890796	0.410186	NaN	NaN	NaN
min	1.000000	1.000000	NaN	NaN	NaN
25%	477.250000	1.000000	NaN	NaN	NaN
50%	926.000000	1.000000	NaN	NaN	NaN
75%	1393.000000	1.000000	NaN	NaN	NaN
max	1834.000000	15.000000	NaN	NaN	NaN

Cột order_id:

- Số lượng đơn hàng: 4622.
- Giá trị nhỏ nhất là 1, lớn nhất là 1834. Có thể suy ra rằng mỗi order_id đại diện cho một đơn hàng cụ thể.
- Giá trị trung bình là 927.25 với độ lệch chuẩn 528.89, cho thấy các đơn hàng được phân bố khá rộng.

Cột quantity:

- Số lượng sản phẩm trong một mục dao động từ 1 đến 15.
- Phần lớn các đơn hàng chỉ có số lượng bằng 1 (thể hiện ở giá trị trung vị, 25%, 50%, và 75%).

Cột item_name:

- Có 50 loại sản phẩm khác nhau.
- Món phổ biến nhất là "Chicken Bowl", xuất hiện 726 lần.

Cột choice_description: Có 1043 lựa chọn khác nhau, nhưng có nhiều giá trị thiếu (NaN), chỉ 3376 trên 4622 mục là có giá trị.

Cột item_price:

- Giá dao động từ \$1.09 đến \$44.25.
- Giá trị phổ biến nhất là \$8.75, xuất hiện 730 lần.
- Giá trung bình là \$7.08 với độ lệch chuẩn \$4.66.

Phát hiện và giải pháp:

- + Cột 3 [choice_description] có 3376 dữ liệu không bị NULL → 1246 giá trị NULL → làm sạch dữ liệu NULL bằng cách chuyển đổi qua chuỗi "none" để tránh ảnh hưởng dữ liệu
- + Không có giá trị = "0" hay NULL trong cột [quantity]
- + Không có dòng nào thiếu tên món ăn trong cột [item_name]
- + Ở cột [item_price]
 - Lúc đầu: đây là dạng object → không rõ kiểu dữ liệu nào để chuyển đổi về dạng số được
 - Lúc sau: chuyển về kiểu dữ liệu string → bỏ dấu \$ → chuyển đổi kiểu dữ liệu thành dạng float (số)



Kiểm tra missing data

- Dữ liệu thiếu:
 - + Cột choice_description có 3376 giá trị không rỗng (khoảng 72% dữ liệu đầy đủ) → làm sạch dữ liệu NULL bằng cách chuyển đổi qua chuỗi “none” để tránh ảnh hưởng dữ liệu

before

```
0      NaN
1  [Clementine]
2    [Apple]
3      NaN
4  [Tomatillo-Red Chili Salsa (Hot), [Black Beans...
Name: choice_description, dtype: object
```

after

```
0      none
1  [Clementine]
2    [Apple]
3      none
4  [Tomatillo-Red Chili Salsa (Hot), [Black Beans...
Name: choice_description, dtype: object
```

- Dữ liệu cần xử lý:

- + Cột `item_price` cần được chuyển đổi từ dạng chuỗi (string) sang kiểu số (float) để phân tích.

```
0      $2.39
1      $3.39
2      $3.39
3      $2.39
4     $16.98
...
4617   $11.75
4618   $11.75
4619   $11.25
4620    $8.75
4621    $8.75
Name: item_price, Length: 4622, dtype: object
```

before

```
0      2.39
1      3.39
2      3.39
3      2.39
4     16.98
...
4617   11.75
4618   11.75
4619   11.25
4620    8.75
4621    8.75
Name: item_price, Length: 4622, dtype: float64
```

after

- **Chuẩn hóa dữ liệu trong ma trận**

```
transaction_df = transaction_df.astype(int)
```

```
transaction_df.values
```

```
array([[0, 0, 0, ..., 0, 0, 0],
       [0, 0, 0, ..., 0, 0, 0],
       [0, 0, 0, ..., 0, 0, 0],
       ...,
       [0, 0, 0, ..., 0, 0, 0],
       [0, 0, 0, ..., 0, 0, 0],
       [0, 0, 0, ..., 0, 0, 0]])
```

III. Tổng hợp dữ liệu

1. Nhóm dữ liệu theo `order_id`, tạo danh sách các sản phẩm trong từng đơn hàng.
Sử dụng hàm `group by`

```
order_id      item_name
0           1  [Chips and Fresh Tomato Salsa, Izze, Nantucket...
1           2                      [Chicken Bowl]
2           3          [Chicken Bowl, Side of Chips]
3           4      [Steak Burrito, Steak Soft Tacos]
4           5      [Steak Burrito, Chips and Guacamole]
...          ...
1829        1830      [Steak Burrito, Veggie Burrito]
1830        1831      [Carnitas Bowl, Chips, Bottled Water]
1831        1832      [Chicken Soft Tacos, Chips and Guacamole]
1832        1833      [Steak Burrito, Steak Burrito]
1833        1834  [Chicken Salad Bowl, Chicken Salad Bowl, Chick...

[1834 rows x 2 columns]
```

2. Tạo ma trận mua chéo

Mục tiêu: Biểu diễn sự đồng xuất hiện của các sản phẩm trong các giao dịch.

Thao tác:

- a. Sử dụng thư viện **mlxtend**:
 - i. **TransactionEncoder** để mã hóa danh sách sản phẩm thành ma trận nhị phân (binary matrix).
- b. Mỗi cột là một sản phẩm, mỗi hàng là một đơn hàng:
 - i. 1: Sản phẩm có trong đơn.
 - ii. 0: Sản phẩm không có trong đơn.

→ Số lượng giao dịch: 1834

Danh sách sản phẩm: ['6 Pack Soft Drink', 'Barbacoa Bowl', 'Barbacoa Burrito', 'Barbacoa Crispy Tacos', 'Barbacoa Salad Bowl', 'Barbacoa Soft Tacos', 'Bottled Water', 'Bowl', 'Burrito', 'Canned Soda', 'Canned Soft Drink', 'Carnitas Bowl', 'Carnitas Burrito', 'Carnitas Crispy Tacos', 'Carnitas Salad', 'Carnitas Salad Bowl', 'Carnitas Soft Tacos', 'Chicken Bowl', 'Chicken Burrito', 'Chicken Crispy Tacos', 'Chicken Salad', 'Chicken Salad Bowl', 'Chicken Soft Tacos', 'Chips', 'Chips and Fresh Tomato Salsa', 'Chips and Guacamole', 'Chips and Mild Fresh Tomato Salsa', 'Chips and Roasted Chili Corn Salsa', 'Chips and Roasted Chili-Corn Salsa', 'Chips and Tomatillo Green Chili Salsa', 'Chips and Tomatillo Red Chili Salsa', 'Chips and Tomatillo-Green Chili Salsa', 'Chips and Tomatillo-Red Chili Salsa', 'Crispy Tacos', 'Izze', 'Nantucket Nectar', 'Salad', 'Side of Chips', 'Steak Bowl', 'Steak Burrito', 'Steak Crispy Tacos', 'Steak Salad',

'Steak Salad Bowl', 'Steak Soft Tacos', 'Veggie Bowl', 'Veggie Burrito', 'Veggie Crispy Tacos', 'Veggie Salad', 'Veggie Salad Bowl', 'Veggie Soft Tacos']

Số lượng sản phẩm: 50

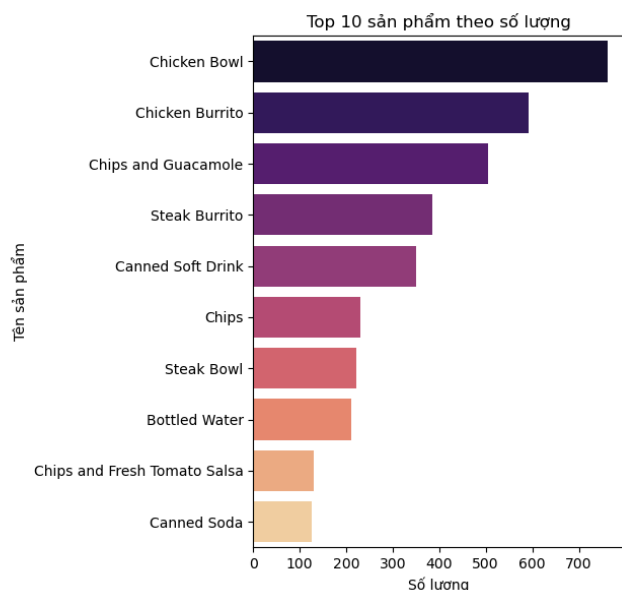
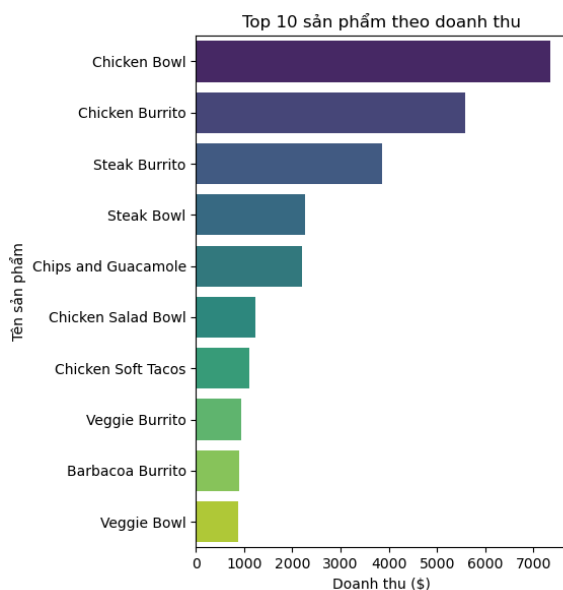
	6 Pack Soft Drink	Barbacoa Bowl	Barbacoa Burrito	Barbacoa Crispy Tacos	\
0	False	False	False	False	False
1	False	False	False	False	False
2	False	False	False	False	False
3	False	False	False	False	False
4	False	False	False	False	False

	Barbacoa Salad Bowl	Barbacoa Soft Tacos	Bottled Water	Bowl	Burrito	\
0	False	False	False	False	False	False
1	False	False	False	False	False	False
2	False	False	False	False	False	False
3	False	False	False	False	False	False
4	False	False	False	False	False	False

	Canned Soda	...	Steak Crispy Tacos	Steak Salad	Steak Salad Bowl	\
0	False	...	False	False	False	False
1	False	...	False	False	False	False
2	False	...	False	False	False	False
3	False	...	False	False	False	False
4	False	...	False	False	False	False

3. Tính tổng số lượng và doanh thu cho từng sản phẩm

→ Ta rút ngắn dữ liệu theo 2 biểu đồ sau:



Nhận xét: 1. Biểu đồ bên trái: Top 10 sản phẩm theo doanh thu

Những sản phẩm mang lại doanh thu cao nhất:

- **Chicken Bowl (7000 USD)** và **Chicken Burrito (~5200 USD)** chiếm các vị trí đầu bảng:
 - Đây là hai sản phẩm chính với giá trị trung bình cao hơn so với các sản phẩm còn lại.

- Doanh thu cao có thể đến từ việc kết hợp giữa giá bán và số lượng tiêu thụ.
- Đây có thể là sản phẩm "cốt lõi" trong thực đơn, được khách hàng ưa chuộng nhất.
- **Steak Bowl (~3600 USD):** Xếp vị trí thứ ba, kém Chicken Burrito một khoảng cách đáng kể, có thể do giá cao hơn hoặc ít phổ biến hơn.

Những sản phẩm phụ đóng góp đáng kể:

- **Chips and Guacamole (~2200 USD):**
 - Là món ăn phụ nhưng vẫn có doanh thu cao, chứng tỏ sản phẩm này được khách hàng mua kèm rất nhiều.
 - Giá trị combo hoặc gói khuyến mãi kèm món chính có thể là chiến lược đáng sau.
- Các món khác như **Veggie Burrito**, **Barbacoa Burrito**, và **Veggie Bowl**:
 - Tuy không mang lại doanh thu cao nhưng vẫn đủ để lọt vào top 10, có thể hướng tới nhóm khách hàng ăn chay hoặc có nhu cầu đặc biệt.

Nhận xét tổng quan:

- Doanh thu tập trung nhiều vào các món chính như **Chicken Bowl** và **Chicken Burrito**, chiếm phần lớn tổng doanh thu của top 10.
- Các món ăn phụ và nước giải khát đóng vai trò hỗ trợ, có giá trị trung bình thấp nhưng hỗ trợ tăng doanh thu tổng thể.

2. Biểu đồ bên phải: Top 10 sản phẩm theo số lượng

Xu hướng tiêu thụ số lượng:

- **Chicken Bowl (~700 lượt mua) và Chicken Burrito (~500 lượt mua):**
 - Hai món này dẫn đầu về số lượng bán ra, tương ứng với vị trí của chúng trong bảng doanh thu.
 - Khách hàng có xu hướng lựa chọn nhiều món này, cho thấy chúng có sức hút mạnh mẽ và phù hợp với khẩu vị phổ thông.
- **Canned Soft Drink (~350 lượt mua) và Chips (~300 lượt mua):**
 - Nước uống đóng hộp và Chips thuộc nhóm sản phẩm có số lượng bán ra cao nhưng không xuất hiện trong biểu đồ doanh thu.
 - Điều này gợi ý rằng giá bán trung bình của chúng thấp, là các sản phẩm hỗ trợ chứ không phải sản phẩm chính.

Những sản phẩm đáng chú ý khác:

- **Steak Burrito và Steak Bowl (~250–300 lượt mua):**

- Được tiêu thụ ít hơn so với các món từ gà (Chicken Bowl/Burrito), nhưng vẫn nằm trong nhóm bán chạy.
- Có thể đây là sản phẩm dành cho phân khúc cao hơn (do giá trung bình cao hơn).
- **Bottled Water và Canned Soda (~150–200 lượt mua):**
 - Là nước uống phổ biến với số lượng tiêu thụ khá lớn, nhưng giá trị đơn hàng thấp dẫn đến không nằm trong top doanh thu.

Nhận xét tổng quan:

- Các món chính vẫn duy trì vị trí hàng đầu, nhưng nhóm sản phẩm giá trị thấp (đồ uống, đồ ăn nhẹ) chiếm phần lớn số lượng tiêu thụ.
- Chiến lược thúc đẩy các món ăn chính kết hợp với món phụ có thể tối ưu hóa cả doanh thu lẫn số lượng.

So sánh giữa doanh thu và số lượng

- **Chicken Bowl:** Dẫn đầu cả về doanh thu và số lượng, chứng tỏ đây là sản phẩm thành công toàn diện.
- **Chips and Guacamole:** Có sự cân bằng tốt, nằm ở giữa trong cả hai biểu đồ, là sản phẩm hỗ trợ hiệu quả.
- **Canned Soft Drink và Chips:** Xuất hiện trong top số lượng nhưng không có mặt trong top doanh thu, phản ánh giá trị đơn hàng thấp.
- **Steak Bowl và Steak Burrito:** Được tiêu thụ ít hơn món từ gà, nhưng vẫn có mức doanh thu tương đối cao nhờ giá bán cao hơn.

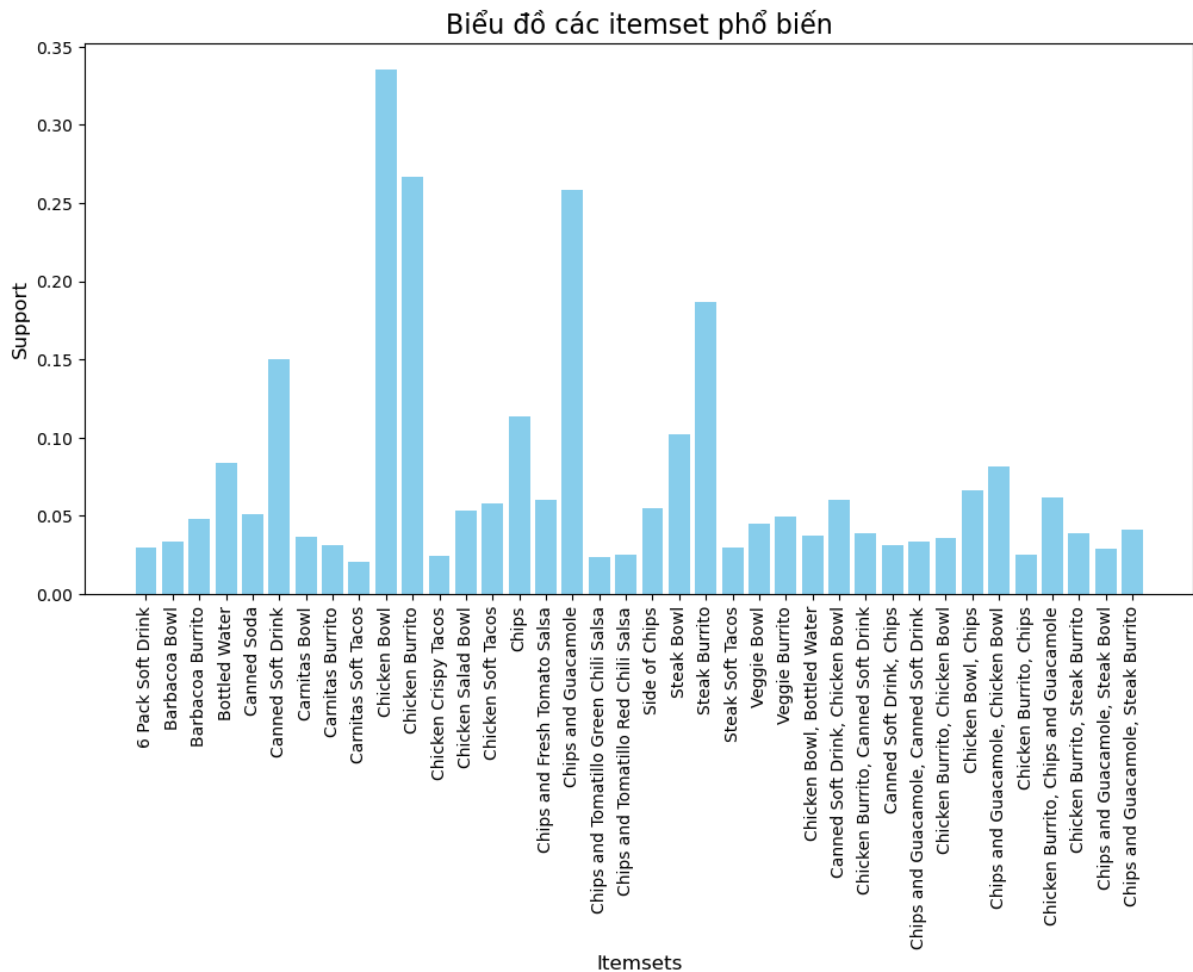
Kết luận và đề xuất

1. **Phát triển thêm các sản phẩm từ gà (Chicken Bowl, Chicken Burrito):**
 - Gà đang là nguyên liệu phổ biến nhất. Đề xuất mở rộng thực đơn với các món sáng tạo dựa trên gà để tận dụng xu hướng này.
2. **Tăng cường bán kèm đồ uống và món phụ:**
 - Các sản phẩm như **Chips**, **Canned Soft Drink**, và **Chips and Guacamole** bán ra với số lượng lớn. Cần có chiến lược upsell/combo để tăng giá trị đơn hàng.
3. **Xem xét quảng bá các món khác:**
 - Các món như **Veggie Burrito** hoặc **Steak Bowl** có tiềm năng nhưng chưa bán ra với số lượng lớn. Có thể cải thiện qua các chương trình khuyến mãi hoặc quảng cáo nhấn mạnh lợi ích của món.

IV. Phân tích dữ liệu

** Khai phá các tập phổ biến (Frequent Itemsets)

- **Mục tiêu:** Tìm các tập hợp sản phẩm thường xuyên được mua cùng nhau.
 1. Áp dụng thuật toán **Apriori**:
 - $\text{min_support} = 0.02$: Ngưỡng tối thiểu để xác định tập phổ biến



Giải thích:

- **Trục hoành (X):** Bao gồm danh sách các **itemsets** (tập hợp các sản phẩm hoặc dịch vụ được mua hoặc sử dụng cùng nhau). Ví dụ: "Chips", "Chicken Bowl", "Chips and Guacamole",...
- **Trục tung (Y):** Thể hiện giá trị **Support**, chỉ số đo tần suất xuất hiện của mỗi itemset trong tập dữ liệu, tính theo tỷ lệ phần trăm (giá trị từ 0 đến 1).
- **Cột màu xanh:** Mỗi cột đại diện cho mức độ phổ biến của một itemset. Cột cao hơn thể hiện itemset được xuất hiện thường xuyên hơn.

- **Tiêu đề:** "Biểu đồ các itemset phổ biến" chỉ rõ đây là phân tích các itemset dựa trên tần suất xuất hiện.

Nhận xét:

- Itemset phổ biến nhất: Itemset có giá trị Support cao nhất là "Chicken Bowl" (khoảng 0.35), cho thấy sản phẩm này được mua thường xuyên nhất trong tập dữ liệu.
- Itemset ít phổ biến hơn: Một số itemsets như "6 Pack Soft Drink" hoặc "Bottled Water" có Support rất thấp, cho thấy chúng ít được mua cùng hoặc có tần suất xuất hiện thấp.
- Các tổ hợp itemsets đặc biệt: "Chips and Guacamole", cũng có giá trị Support khá cao, điều này có thể chỉ ra rằng khách hàng thường mua các sản phẩm này cùng nhau.
- Các itemsets phổ biến như "Chicken Bowl", "Chips and Guacamole" có thể là sản phẩm chủ lực, cần tập trung vào các chiến lược marketing hoặc quảng cáo.
- Với các itemsets có Support thấp, có thể xem xét loại bỏ chúng hoặc cải thiện chiến lược bán hàng để tăng tính phổ biến.

2. Lấy các tập phổ biến và số lần xuất hiện

mục 2 của ptdl: lift đo lường mối quan hệ mạnh mẽ giữa các sản phẩm

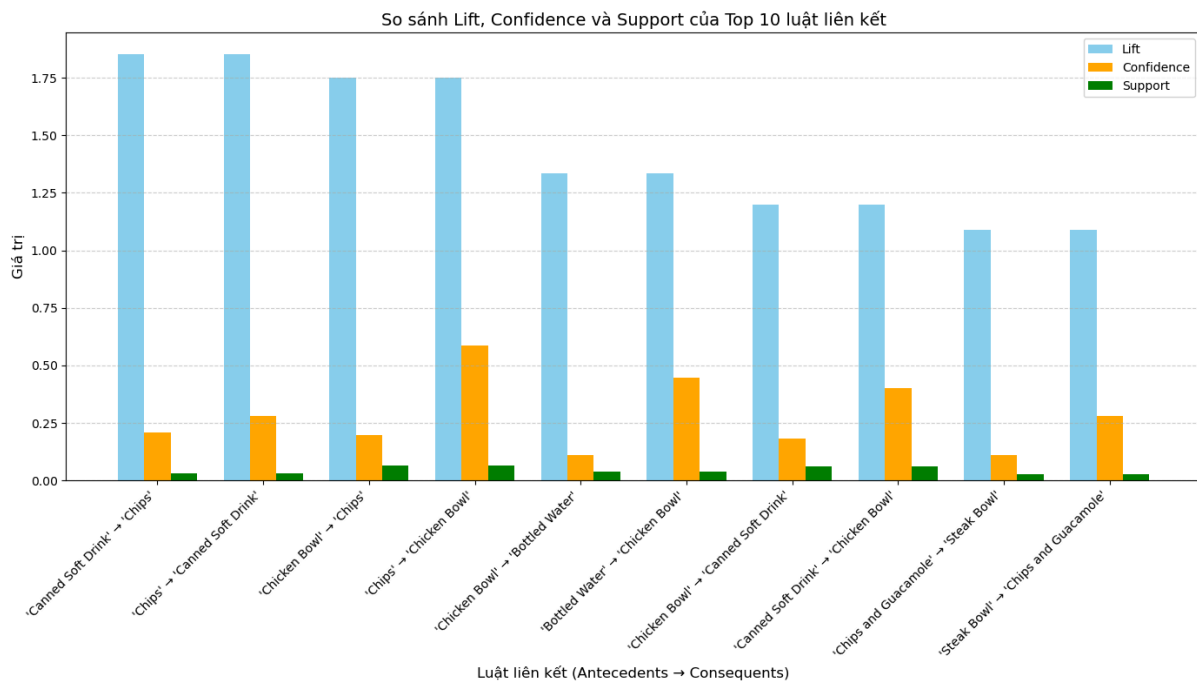
Nếu lift > 1: Sự xuất hiện của các sản phẩm antecedents làm tăng khả năng xuất hiện của consequents.

min_threshold=1.0: Ngưỡng tối thiểu cho giá trị lift. Chỉ giữ lại các quy tắc có lift ≥ 1 (tức là các sản phẩm liên quan đến nhau một cách tích cực).

- antecedents: Tập sản phẩm xuất hiện đầu tiên (bên trái mũi tên trong quy tắc liên kết).
- consequents: Tập sản phẩm xuất hiện sau (bên phải mũi tên trong quy tắc liên kết).
- support: Tỷ lệ giao dịch chứa cả antecedents và consequents.
- confidence: Xác suất xảy ra của consequents khi antecedents đã xảy ra.

lift: Độ mạnh của mối quan hệ giữa antecedents và consequents. Giá trị lift càng cao, mối quan hệ càng chặt chẽ.

- Khi mua chips, có 50% khả năng mua chicken bowl.
- Mối liên hệ giữa chips và chicken bowl mạnh hơn ngẫu nhiên (lift = 1.75).



Nhận xét:

1. Luật có Lift cao:

- "Canned Soft Drink → Chips" và "Chips → Canned Soft Drink" có Lift cao nhất (~1.75), cho thấy sự liên kết mạnh giữa Canned Soft Drink và món phụ Chips.
- Các luật có Chicken Bowl, Bottled Water và Chips cũng đạt Lift cao, chứng minh sự phổ biến của các món này trong tổ hợp.

2. Confidence (Độ tin cậy):

- Luật "Canned Soft Drink → Chips" và "Chips → Canned Soft Drink" có Confidence cao nhất (>50%), thể hiện khả năng khách hàng mua món này kèm nhau rất cao.
- Các luật có Chicken Bowl và Bottled Water đạt Confidence tốt, đặc biệt khi đi kèm Chips.

3. Support (Hỗ trợ):

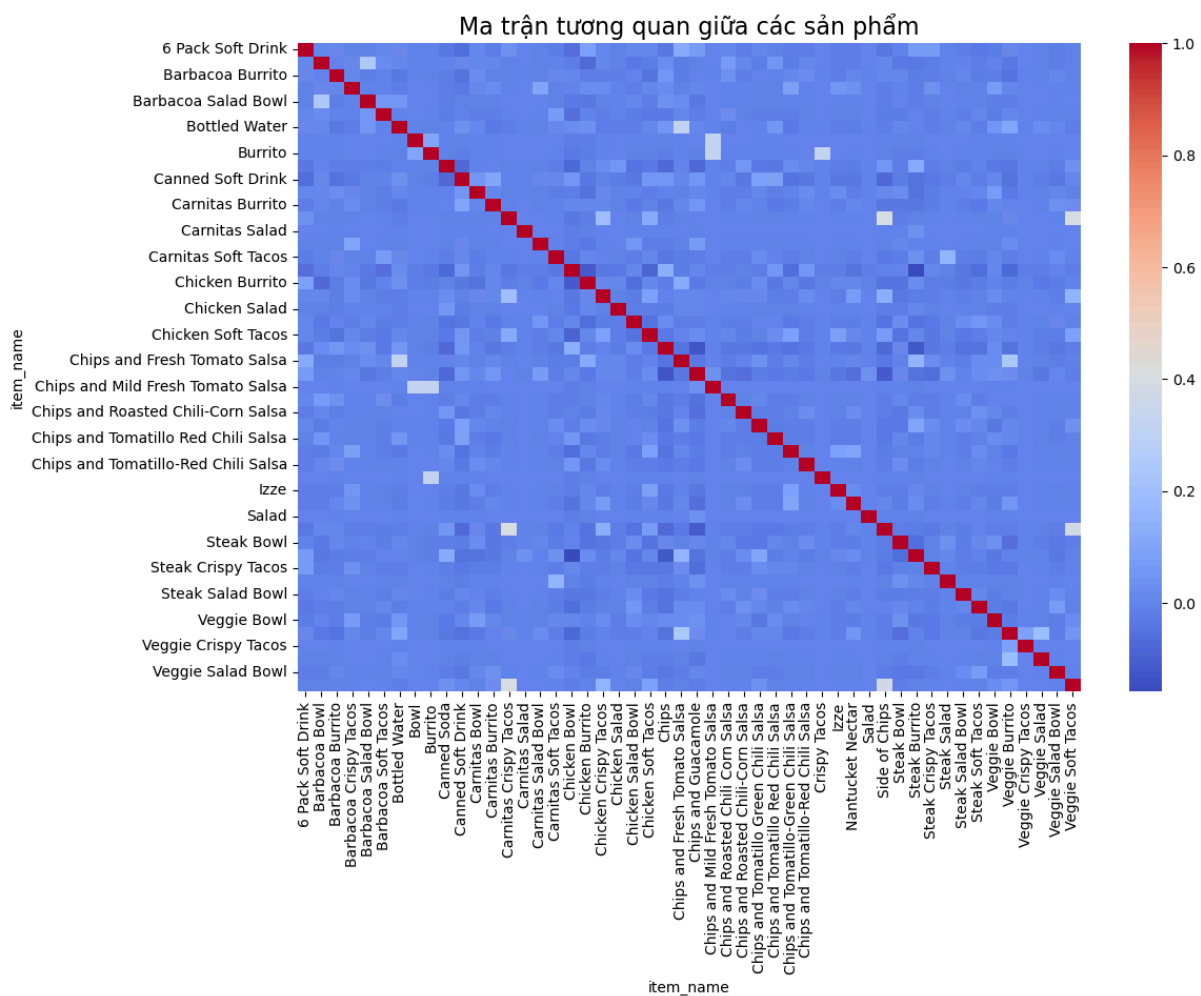
- Support thấp ở tất cả các luật, cho thấy các luật liên kết này không xuất hiện quá thường xuyên trong toàn bộ dữ liệu giao dịch. Tuy nhiên, giá trị Support tăng dần với các món phổ biến như Chips, Canned Soft Drink và Chicken Bowl.

Tóm lại:

- Canned Soft Drink và Chips là cặp món liên kết mạnh mẽ nhất.
- Chicken Bowl là món chính phổ biến, có liên kết tốt với cả Chips và đồ uống.
- Các luật với Bottled Water cũng đáng chú ý, đặc biệt cho nhóm khách hàng quan tâm đến sức khỏe.

→ Đẩy mạnh bán chéo các món phổ biến với luật liên kết có Lift cao.

3. Tạo danh sách sản phẩm trong từng đơn hàng



Ý nghĩa:

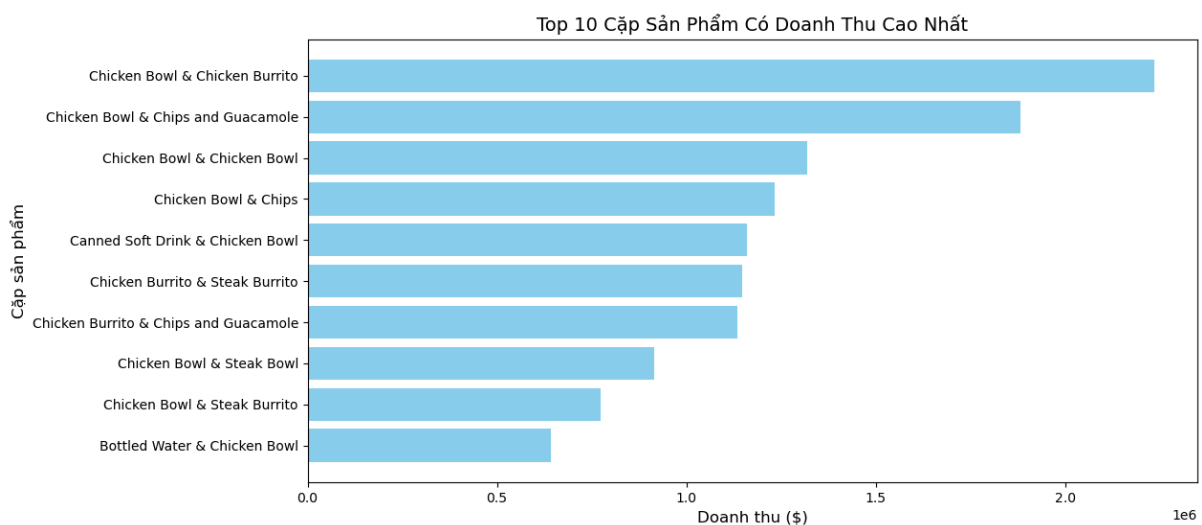
- **Màu đỏ (gần giá trị 1):** Biểu thị mối tương quan dương mạnh, nghĩa là hai sản phẩm này thường xuyên được mua cùng nhau.
- **Màu xanh nhạt hoặc trung tính (gần giá trị 0):** Mối tương quan yếu hoặc không có sự liên quan đáng kể giữa hai sản phẩm.

- **Màu đỏ đậm trên đường chéo chính:** Hệ số tương quan của một sản phẩm với chính nó (luôn bằng 1).
- **Đường chéo chính:** Đường chéo chính có màu đỏ đậm, vì đây là mối tương quan của mỗi sản phẩm với chính nó (luôn bằng 1). Điều này là hợp lý trong mọi ma trận tương quan.
- **Các sản phẩm có mối tương quan cao:** Một số ô gần màu đỏ cho thấy có những sản phẩm thường được mua cùng nhau. Ví dụ:
 - + **"Chips and Fresh Tomato Salsa" và "Chips and Mild Fresh Tomato Salsa":** Có thể được mua cùng nhau do là các món ăn kèm phổ biến.
 - + Các sản phẩm tương tự hoặc bổ sung nhau (chips và salsa) có xu hướng tương quan cao.
- **Các sản phẩm ít tương quan:** Phần lớn các ô có màu xanh, cho thấy đa số sản phẩm trong tập dữ liệu có mối tương quan thấp. Điều này cho thấy người dùng thường không mua tất cả các sản phẩm cùng nhau mà chỉ mua theo nhu cầu cụ thể.
- **Sự phổ biến của sản phẩm:** Nếu một sản phẩm thường xuyên xuất hiện với các sản phẩm khác, nó sẽ có nhiều ô màu đỏ hoặc nhạt dần xung quanh hàng và cột của nó. Ví dụ, các món ăn phổ biến như "Chicken Burrito" hoặc "Chips" có thể xuất hiện với nhiều sản phẩm khác.

Nhận xét:

- Biểu đồ này rất hữu ích để phân tích hành vi mua hàng và nhận biết các sản phẩm thường được mua cùng nhau.
- Có một số sản phẩm có mối tương quan cao, đặc biệt là những sản phẩm thuộc cùng danh mục (ví dụ, món ăn kèm hoặc đồ uống).

4. Tạo doanh thu theo cặp sản phẩm



Nhận xét:

- **Sản phẩm có doanh thu cao nhất:** Cặp sản phẩm "**Chicken Bowl & Chicken Burrito & Chicken Burrito**" đứng đầu với doanh thu vượt trội (trên 2 triệu USD). Điều này cho thấy đây là nhóm sản phẩm chủ lực và rất được ưa chuộng.
- **Các cặp sản phẩm phổ biến tiếp theo:**
 - Cặp sản phẩm "**Chicken Bowl & Chips and Guacamole & Chips and Guacamole**" có doanh thu xếp thứ hai, đạt gần mức cao nhất.
 - Những cặp sản phẩm như "**Chicken Bowl & Chicken Bowl & Chicken Bowl**" và "**Chicken Bowl & Chips & Chips**" cũng có doanh thu đáng kể, cho thấy sự phổ biến của "Chicken Bowl" và "Chips" trong danh mục.
- **Tính đa dạng của các cặp sản phẩm:**
 - Các tổ hợp sản phẩm phổ biến có xu hướng kết hợp các món chính (như "Chicken Bowl", "Chicken Burrito") với các món phụ hoặc đồ ăn kèm (như "Chips", "Guacamole").
 - Một số tổ hợp đồ uống (ví dụ: "Canned Soft Drink") xuất hiện nhưng có doanh thu thấp hơn so với các cặp thực phẩm.
- **Tầm quan trọng của món chính:** "Chicken Bowl" và "Chicken Burrito" xuất hiện nhiều lần trong các cặp có doanh thu cao, chứng tỏ đây là các món chính mang lại giá trị lớn cho doanh nghiệp.
- **Doanh thu giảm dần theo thứ hạng:**
 - Biểu đồ cho thấy sự giảm dần doanh thu từ cặp sản phẩm cao nhất đến thấp nhất. Tuy nhiên, sự chênh lệch giữa các cặp đứng đầu và các cặp cuối trong danh sách là không quá lớn, cho thấy sự đóng góp đồng đều của nhiều nhóm sản phẩm.

Gợi ý:

- **Tập trung vào món chính:** Tăng cường quảng bá các sản phẩm chủ lực như "Chicken Bowl", "Chicken Burrito", và các món ăn kèm như "Chips" và "Guacamole".
- **Chiến lược upsell và combo:** Tạo các combo kết hợp món chính với món phụ và đồ uống để tối ưu hóa doanh thu.
- **Tăng trải nghiệm khách hàng:** Đẩy mạnh dịch vụ với các tổ hợp sản phẩm phổ biến để giữ chân khách hàng và thúc đẩy doanh số.

5. Tạo dữ liệu dự đoán, huấn luyện mô hình Logistic Regression

- **Mục đích:** Tạo dữ liệu dự đoán, chú ý tới những dữ liệu riêng lẻ có lượt bán và doanh thu cao, huấn luyện mô hình, thử nghiệm xem sản phẩm đó có hay được

mua chung với các sản phẩm khác cũng có doanh thu hoặc số lượng bán ra cao hay không

- Giải thích:

- + Precision: Tỷ lệ dự đoán đúng trong số các mẫu được dự đoán là thuộc về một lớp cụ thể.
- + Recall: Tỷ lệ dự đoán đúng trong số các mẫu thực sự thuộc về lớp cụ thể đó.
- + F1-Score: Trung bình điều hòa của precision và recall, cân bằng giữa độ chính xác và độ bao phủ.
- + Support: Số lượng mẫu thực sự thuộc về từng lớp.
- + Lớp 0: % sản phẩm không được mua
- + Lớp 1: % sản phẩm được mua
- + VIF<5: không có hiện tượng đa cộng tuyến ->KHÔNG LOẠI BỎ BIẾN NÀO KHI CÓ HỆ SỐ ÂM

- Sau khi thử nghiệm, nhận thấy sự thiên vị giữa lớp 0 và lớp 1, cần áp dụng SMOTE để tăng dữ liệu lớp 1 → Mô hình có hiệu suất tốt, đặc biệt trong việc cân bằng giữa hai lớp. Precision và Recall đều ở mức ổn định, và không có sự thiên lệch rõ rệt.

- Các dữ liệu được chọn để dự đoán

- Dựa vào top 10 tổng số lượng hoặc doanh thu có lượt bán ra cao nhất, ta sẽ lựa chọn các sản phẩm: Chicken bowl, chicken burrito, chips and guacamole, steak burrito, steak bowl, canned soft drink, chips, bottled water, chips and fresh tomato salsa, canned soda, Chicken Salad Bowl, Chicken Soft Tacos, Veggie Burrito, Veggie Bowl, Barbacoa Burrito → đây là những biến dự đoán
- Xác định 15 sản phẩm có tác động tới biến dự đoán
 - a. Chicken bowl

Đánh giá mô hình:

	precision	recall	f1-score	support
0	0.79	0.83	0.81	243
1	0.82	0.79	0.80	245
accuracy			0.81	488
macro avg	0.81	0.81	0.81	488
weighted avg	0.81	0.81	0.81	488

Độ chính xác: 0.8073770491803278

- + Precision của lớp 1 (0.82) cao hơn Recall (0.79), cho thấy mô hình ưu tiên dự đoán chính xác
- + Mô hình hoạt động khá đồng đều giữa lớp 0 và lớp 1 với F1-Score lần lượt là 0.81 và 0.80. Điều này cho thấy mô hình không bị thiên lệch quá mức về một lớp.
- + Với Accuracy là 81%, mô hình đạt hiệu suất tốt, đặc biệt khi dữ liệu giữa hai lớp cân bằng.

- + Lớp 1:
 - + Precision: 0.82: Trong tất cả các mẫu được dự đoán là lớp 1, có 82% là dự đoán đúng.
 - + Recall: 0.79: Trong tất cả các mẫu thực tế thuộc lớp 1, mô hình nhận diện đúng 79%.
 - + F1-Score: 0.80: Kết quả này cho thấy lớp 1 có hiệu suất tương đương lớp 0, mặc dù Recall thấp hơn một chút.

b. Chicken Burrito

Đánh giá mô hình:

	precision	recall	f1-score	support
0	0.81	0.85	0.83	285
1	0.82	0.78	0.80	253
accuracy			0.82	538
macro avg	0.82	0.82	0.82	538
weighted avg	0.82	0.82	0.82	538

Độ chính xác: 0.8178438661710037

- Lớp 0:
 - Precision: 0.8: Trong các dự đoán là lớp 0, có 81% là đúng.
 - Recall: 0.85: Trong các mẫu thực tế thuộc lớp 0, mô hình nhận diện đúng 85%.
 - F1-Score: 0.83: Cân bằng giữa Precision và Recall, cho thấy hiệu suất tốt ở lớp này.
- Lớp 1:
 - Precision: 0.82: Trong các dự đoán là lớp 1, có 82% là đúng.
 - Recall: 0.78: Trong các mẫu thực tế thuộc lớp 1, mô hình nhận diện đúng 78%.
 - F1-Score: 0.80: Hiệu suất của lớp này hơi thấp hơn so với lớp 0, đặc biệt là ở Recall.

Tổng quan hiệu suất:

- Accuracy: 0.82: Mô hình dự đoán đúng 82% tổng số mẫu.
- Macro Avg (trung bình đơn giản): 0.82: Trung bình cộng của các chỉ số Precision, Recall và F1-Score giữa các lớp.
- Weighted Avg (trung bình trọng số): 0.82: Trung bình có trọng số, tính đến số lượng mẫu ở từng lớp. Vì dữ liệu khá cân bằng, Weighted Avg gần như bằng Macro Avg.

c. Chips and Guacamole

```
Đánh giá mô hình:
      precision    recall  f1-score   support

0         0.84        0.69        0.75        280
1         0.72        0.86        0.78        264

 accuracy          0.77        544
 macro avg         0.78        0.77        0.77        544
 weighted avg         0.78        0.77        0.77        544

Độ chính xác: 0.7702205882352942
```

Precision:

- Lớp "1" (Chips and Guacamole được chọn): Precision đạt 0.72, nghĩa là 72% các dự đoán của mô hình về lớp này là chính xác.
- Lớp "0" đạt 0.84, cho thấy độ chính xác cao hơn khi dự đoán lớp này.

Recall:

- Lớp "1" có recall là 0.86, biểu thị mô hình tìm được 86% các trường hợp thực sự thuộc lớp này.
- Lớp "0" chỉ đạt 0.69, nghĩa là mô hình bỏ sót một lượng đáng kể các trường hợp của lớp này.

F1-score:

- F1-score của lớp "1" đạt 0.78, là mức khá tốt, cân bằng giữa precision và recall.
- Trung bình có trọng số (weighted avg): F1-score là 0.77, biểu thị hiệu năng tổng thể ổn định.

Accuracy: Độ chính xác của mô hình là 0.77, nghĩa là 77% các dự đoán đúng với nhãn thực tế.

Macro Avg: macro avg cho F1-score là 0.77, cho thấy mức hiệu quả tương tự giữa hai lớp.

→ Mô hình logistic regression thể hiện hiệu suất khá tốt, đặc biệt là đối với biến mục tiêu Chips and Guacamole (lớp "1"). Precision và recall cao cho lớp này cho thấy mô hình có thể dự đoán đúng mục tiêu một cách đáng tin cậy.

d. Steak Burrito

```
Đánh giá mô hình:
      precision    recall  f1-score   support

     0       0.90      0.85      0.87        318
     1       0.84      0.89      0.86        279

 accuracy              0.87        597
 macro avg       0.87      0.87      0.87        597
 weighted avg    0.87      0.87      0.87        597

Độ chính xác: 0.8693467336683417
```

Accuracy (Độ chính xác): 0.87 (87%), một con số khá tốt, cho thấy mô hình phân loại hoạt động hiệu quả.

Precision: Lớn hơn 0.80 cho cả hai lớp (0 và 1), nghĩa là mô hình ít dự đoán sai.

Recall: Tương đối cân bằng (0.85 cho lớp 0 và 0.89 cho lớp 1), chỉ ra rằng mô hình dự đoán tốt trên cả hai lớp.

F1-Score: Đạt 0.87 cho cả hai lớp, là thước đo tổng hợp giữa precision và recall, cho thấy mô hình duy trì sự cân bằng giữa độ chính xác và khả năng bao phủ.

→ Mô hình đạt độ chính xác cao (87%), với các chỉ số precision, recall và F1-score đều ở mức cân bằng và tốt. Điều này cho thấy mô hình hoạt động hiệu quả trong việc phân loại dữ liệu và có thể ứng dụng vào thực tế.

e. Steak Bowl

```
Đánh giá mô hình:
      precision    recall  f1-score   support

     0       0.88      0.84      0.86        339
     1       0.84      0.88      0.86        320

 accuracy              0.86        659
 macro avg       0.86      0.86      0.86        659
 weighted avg    0.86      0.86      0.86        659

Độ chính xác: 0.8603945371775418
```

Precision: Độ chính xác khi dự đoán đúng (0.88 cho nhãn 0, 0.84 cho nhãn 1).

Recall: Khả năng tìm đúng các mẫu (0.84 cho nhãn 0, 0.88 cho nhãn 1).

F1-score: Cân bằng giữa Precision và Recall (đều 0.86).

Accuracy: 86%, mô hình dự đoán tốt.

Macro avg: Trung bình giữa các nhãn, Weighted avg: Có trọng số theo số mẫu.

→ Mô hình đạt độ chính xác khá cao (86%) với độ cân bằng tốt giữa các chỉ số Precision và Recall.

f. Canned Soft Drink

Đánh giá mô hình:				
	precision	recall	f1-score	support
0	0.88	0.68	0.77	337
1	0.70	0.89	0.79	287
accuracy			0.78	624
macro avg	0.79	0.78	0.78	624
weighted avg	0.80	0.78	0.77	624

Độ chính xác: 0.7756410256410257

- Precision: Độ chính xác khi dự đoán đúng là 0.88 (nhãn 0) và 0.70 (nhãn 1).
- Recall: Khả năng tìm đúng mẫu là 0.68 (nhãn 0) và 0.89 (nhãn 1).
- F1-score: Cân bằng giữa Precision và Recall, đạt 0.77 (nhãn 0) và 0.78 (nhãn 1).
- Accuracy: 78.6%, mô hình dự đoán ở mức trung bình.

→ Mô hình có độ chính xác khá tốt nhưng không cân bằng giữa Precision và Recall. Với nhãn 1 (Canned Soft Drink), Recall cao hơn Precision, cho thấy mô hình có xu hướng nhận diện nhãn này nhưng đôi khi dự đoán sai.

g. Chips

Đánh giá mô hình:				
	precision	recall	f1-score	support
0	0.90	0.77	0.83	344
1	0.78	0.91	0.84	307
accuracy			0.84	651
macro avg	0.84	0.84	0.84	651
weighted avg	0.85	0.84	0.84	651

Độ chính xác: 0.8356374807987711

Accuracy: Mô hình dự đoán chính xác 84% trường hợp.

Precision:

- Lớp 0: 0.90 (Cao, dự đoán lớp 0 rất đáng tin cậy).
- Lớp 1: 0.78 (Thấp hơn, dự đoán lớp 1 dễ bị nhầm lẫn).

Recall:

- Lớp 0: 0.77 (Mô hình phát hiện tốt nhưng bỏ sót khoảng 23% trường hợp thực sự là lớp 0).
- Lớp 1: 0.91 (Rất cao, mô hình nhận diện gần như toàn bộ lớp 1).

F1-score:

- Lớp 0: 0.83 (Tốt, nhưng cần cải thiện Recall).
- Lớp 1: 0.84 (Ổn định, cân bằng giữa Precision và Recall).

Macro Avg: 0.84 (Hiệu suất tổng thể cân bằng giữa hai lớp).

Weighted Avg: 0.85 (Hiệu suất mô hình ưu tiên lớp có nhiều mẫu hơn).

→ Mô hình hoạt động hiệu quả với độ chính xác cao, cân bằng giữa Precision và Recall.

h. Bottled Water

Đánh giá mô hình:

	precision	recall	f1-score	support
0	0.85	0.77	0.81	356
1	0.77	0.84	0.80	316
accuracy			0.81	672
macro avg	0.81	0.81	0.81	672
weighted avg	0.81	0.81	0.81	672

Độ chính xác: 0.8050595238095238

Accuracy: 81%, mô hình có độ chính xác tốt.

Precision: Nhóm 0 (85%) cao hơn nhóm 1 (77%), mô hình ít dự đoán sai khi khách không chọn Bottled Water.

Recall: Nhóm 1 (84%) cao hơn nhóm 0 (77%), mô hình tốt trong nhận diện khách hàng chọn Bottled Water.

F1-score: Cân đối tốt giữa Precision và Recall, đạt 0.81 ở cả hai nhóm.

i. Chips and Fresh Tomato Salsa

```
Đánh giá mô hình:
      precision    recall  f1-score   support

     0       0.93      0.83      0.88       344
     1       0.85      0.93      0.89       346

 accuracy              0.88       690
 macro avg           0.89      0.88      0.88       690
weighted avg           0.89      0.88      0.88       690

Độ chính xác: 0.8840579710144928
```

Accuracy: 0.88 (~88%), mô hình dự đoán chính xác cao.

Precision: Lớp 0: 0.93, lớp 1: 0.85.

Recall: Lớp 0: 0.83, lớp 1: 0.93.

F1-score: Lớp 0: 0.88, lớp 1: 0.89.

→ Mô hình hoạt động tốt với độ chính xác và hiệu suất cân bằng trên cả hai lớp (~88-89%).

j. Canned Soda

```
Đánh giá mô hình:
      precision    recall  f1-score   support

     0       0.93      0.80      0.86       384
     1       0.79      0.92      0.85       312

 accuracy              0.86       696
 macro avg           0.86      0.86      0.86       696
weighted avg           0.87      0.86      0.86       696

Độ chính xác: 0.8563218390804598
```

Accuracy: Mô hình dự đoán chính xác 86% trường hợp. Đây là một mức hiệu suất khá tốt trong bài toán phân loại này.

Precision:

- Lớp 0: 0.93 (Rất cao). Điều này cho thấy hầu hết các dự đoán là lớp 0 đều chính xác.
- Lớp 1: 0.80 (Thấp hơn). Khoảng 80% các trường hợp được dự đoán là lớp 1 thực sự đúng.

Recall :

- Lớp 0: 0.88 (Cao). Mô hình phát hiện chính xác 88% các trường hợp thực sự là lớp 0.
- Lớp 1: 0.86 (Tốt). Mô hình nhận diện 86% các trường hợp thực sự là lớp 1.

F1-score:

- Lớp 0: 0.90 (Rất tốt). Mô hình hoạt động ổn định với lớp 0.
- Lớp 1: 0.83 (Khá tốt). Mô hình đạt hiệu suất tương đối ở lớp 1, nhưng còn dư địa cải thiện.

→ Mô hình đạt hiệu suất tổng thể cao

k. Chicken Salad Bowl

Đánh giá mô hình:				
	precision	recall	f1-score	support
0	0.89	0.85	0.87	357
1	0.85	0.89	0.87	338
accuracy			0.87	695
macro avg	0.87	0.87	0.87	695
weighted avg	0.87	0.87	0.87	695

Độ chính xác: 0.8705035971223022

Precision, recall và f1-score đều khá cao (~0.87) cho cả hai nhãn.

Độ chính xác tổng thể (accuracy) là 87.3%, cho thấy mô hình dự đoán tốt.

l. Chicken Soft Tacos

Đánh giá mô hình:				
	precision	recall	f1-score	support
0	0.92	0.82	0.86	371
1	0.81	0.91	0.86	320
accuracy			0.86	691
macro avg	0.86	0.87	0.86	691
weighted avg	0.87	0.86	0.86	691

Độ chính xác: 0.8625180897250362

Độ chính xác tổng thể (accuracy): 86.12%

Precision: 0.92 (cho lớp 0) và 0.82 (cho lớp 1).

Recall: 0.85 (cho lớp 0) và 0.86 (cho lớp 1).

F1-score: 0.88 (cho lớp 0) và 0.84 (cho lớp 1).

m. Veggie Burrito

```
Đánh giá mô hình:
      precision    recall  f1-score   support

     0       0.95      0.91      0.93        350
     1       0.91      0.95      0.93        348

 accuracy          0.93          0.93          0.93        698
 macro avg       0.93      0.93      0.93        698
weighted avg       0.93      0.93      0.93        698

Độ chính xác: 0.9297994269340975
```

Độ chính xác (Accuracy): 92.98%

Precision, Recall, F1-score:

- Lớp 0: Precision 0.95, Recall 0.91, F1-score 0.93
- Lớp 1: Precision 0.91, Recall 0.95, F1-score 0.93

Cân bằng: Macro avg và Weighted avg đều đạt 0.93, cho thấy mô hình hoạt động ổn định và cân bằng giữa các lớp.

n. Veggie Bowl

```
Đánh giá mô hình:
      precision    recall  f1-score   support

     0       0.89      0.86      0.87        379
     1       0.84      0.87      0.86        322

 accuracy          0.86          0.86          0.86        701
 macro avg       0.86      0.87      0.86        701
weighted avg       0.87      0.86      0.86        701

Độ chính xác: 0.8644793152639088
```

Độ chính xác (Accuracy): 86.47%

Precision, Recall, F1-score:

- Lớp 0 (không phải Veggie Bowl): Precision 0.87, Recall 0.89, F1-score 0.88
- Lớp 1 (Veggie Bowl): Precision 0.85, Recall 0.82, F1-score 0.84

Cân bằng: Macro avg và Weighted avg đạt 0.86, cho thấy mô hình tương đối cân bằng giữa các lớp.

o. Barbacoa Burrito

```
Đánh giá mô hình:
      precision    recall  f1-score   support

     0       0.92      0.86      0.89        370
     1       0.86      0.92      0.89        329

 accuracy          0.89          699
 macro avg       0.89      0.89      0.89          699
 weighted avg    0.89      0.89      0.89          699

Độ chính xác: 0.8898426323319027
```

Độ chính xác (Accuracy): 88.98%

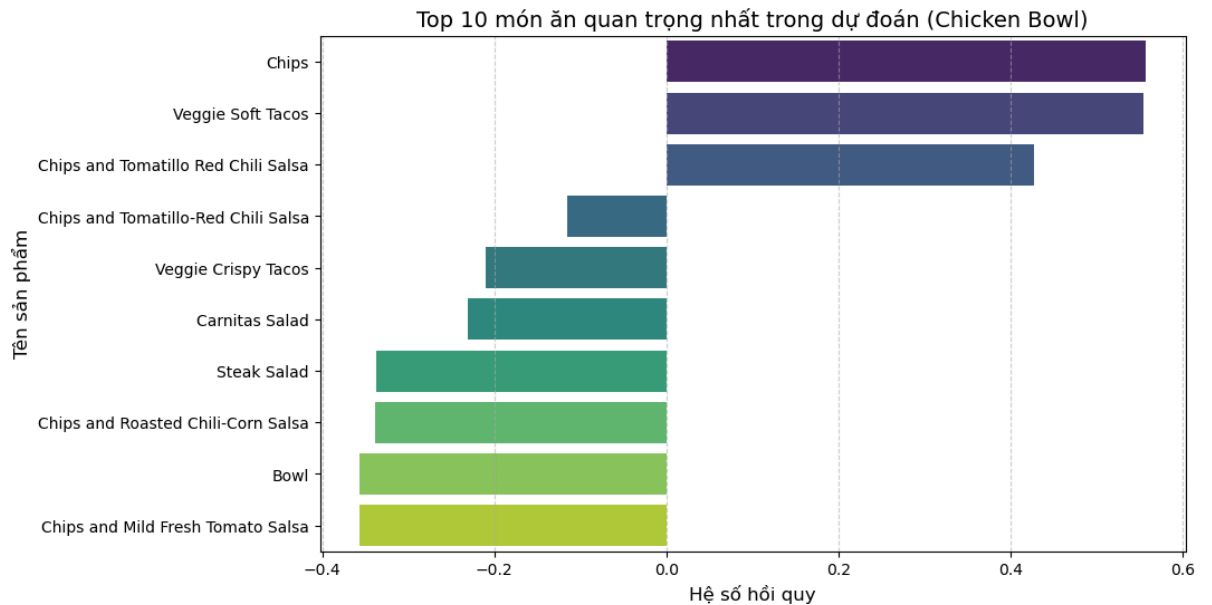
Precision, Recall, F1-score:

- **Lớp 0 (không phải Barbacoa Burrito):** Precision 0.92, Recall 0.86, F1-score 0.89
- **Lớp 1 (Barbacoa Burrito):** Precision 0.86, Recall 0.92, F1-score 0.89

Cân bằng: Macro avg và Weighted avg đều đạt 0.89, chứng tỏ mô hình hoạt động tốt cho cả hai lớp.

V. Trực quan hóa dữ liệu

a. Chicken bowl



Món ăn có ảnh hưởng tích cực lớn nhất:

- Chips: Có hệ số hồi quy cao nhất (~0.55), nghĩa là nếu món này xuất hiện, khả năng khách hàng chọn Chicken Bowl tăng mạnh.
- Veggie Soft Tacos và Chips and Tomatillo Red Chili Salsa cũng có ảnh hưởng tích cực đáng kể.

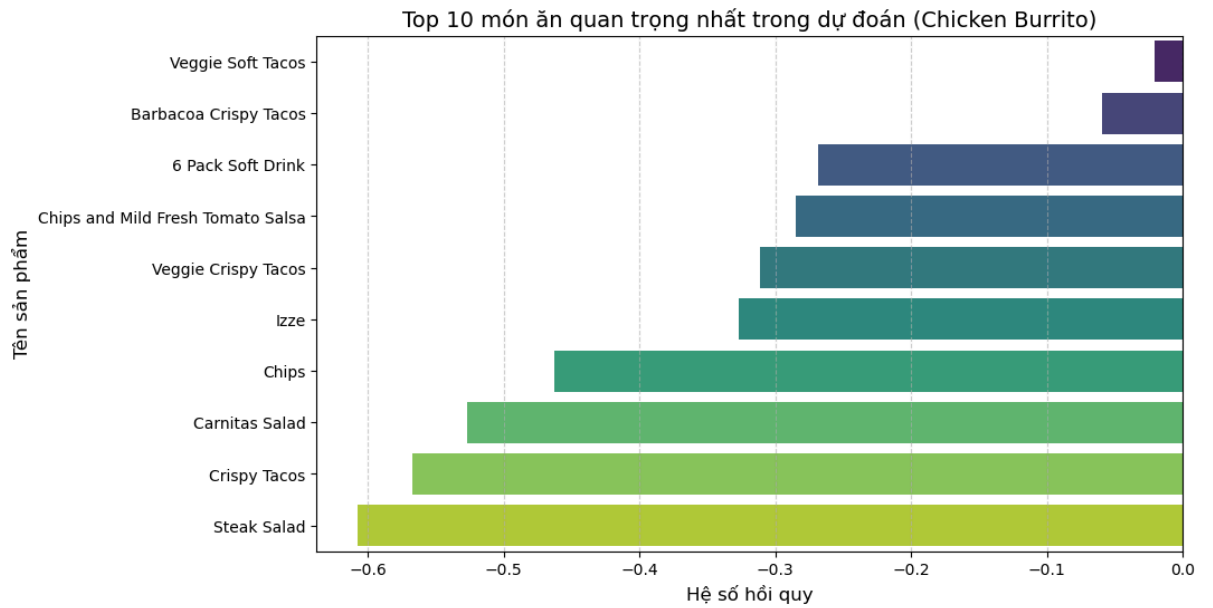
Món ăn có ảnh hưởng tiêu cực lớn nhất:

- Chips and Mild Fresh Tomato Salsa: Ảnh hưởng tiêu cực lớn nhất, với hệ số khoảng -0.35. Sự hiện diện của món này làm giảm khả năng khách hàng chọn Chicken Bowl.
- Các món khác như Bowl và Chips and Roasted Chili-Corn Salsa cũng có ảnh hưởng tiêu cực đáng kể.

Cần nhắc thiết kế thực đơn:

- Các món ăn có hệ số dương (tích cực) như Chips, Veggie Soft Tacos, và Chips and Tomatillo Red Chili Salsa nên được kết hợp quảng bá hoặc bán kèm với Chicken Bowl để tối ưu doanh thu.
- Các món có hệ số âm (tiêu cực) như Chips and Mild Fresh Tomato Salsa và Bowl nên được kiểm tra kỹ lưỡng để tránh làm giảm khả năng khách hàng chọn Chicken Bowl.

b. Chicken Burrito



* Các món tăng khả năng dự đoán Chicken Burrito (hệ số dương lớn):

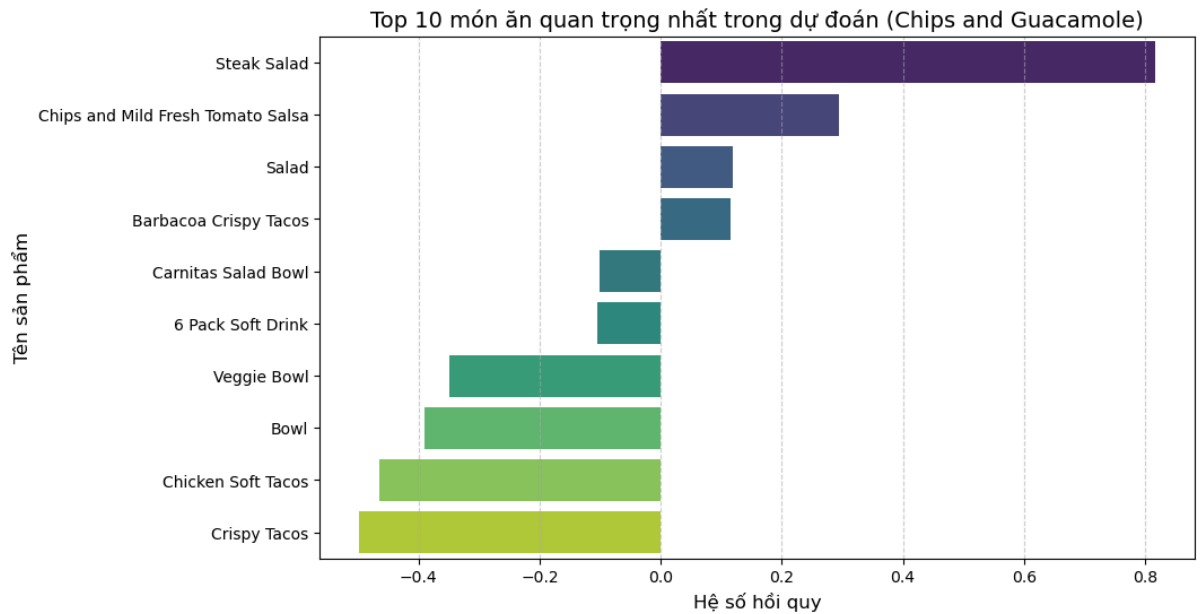
- Veggie Soft Tacos: Ảnh hưởng tích cực cao nhất.
- Barbacoa Crispy Tacos và 6 Pack Soft Drink cũng có ảnh hưởng lớn.

* Các món giảm khả năng dự đoán Chicken Burrito (hệ số âm lớn):

- Steak Salad: Ảnh hưởng tiêu cực mạnh nhất, làm giảm khả năng dự đoán Chicken Burrito.
- Crispy Tacos và Carnitas Salad cũng giảm xác suất dự đoán Chicken Burrito đáng kể.

* Ý nghĩa thực tế: Những món ăn có hệ số hồi quy lớn (dương hoặc âm) là những yếu tố quan trọng cần xem xét khi tối ưu hóa dự đoán. Khách hàng mua Chicken Bowl thường có xu hướng mua thêm Veggie Soft Tacos., trong khi khách hàng chọn Steak Salad ít khả năng mua Chicken Burrito.

c. Chips and Guacamole



Món ăn có ảnh hưởng tích cực lớn nhất:

- **Steak Salad:** Có hệ số hồi quy cao nhất (~ 0.6), nghĩa là nếu món này xuất hiện, khả năng khách hàng chọn Chips and Guacamole tăng mạnh.
- **Chips and Mild Fresh Tomato Salsa** và **Salad** cũng có ảnh hưởng tích cực đáng kể.

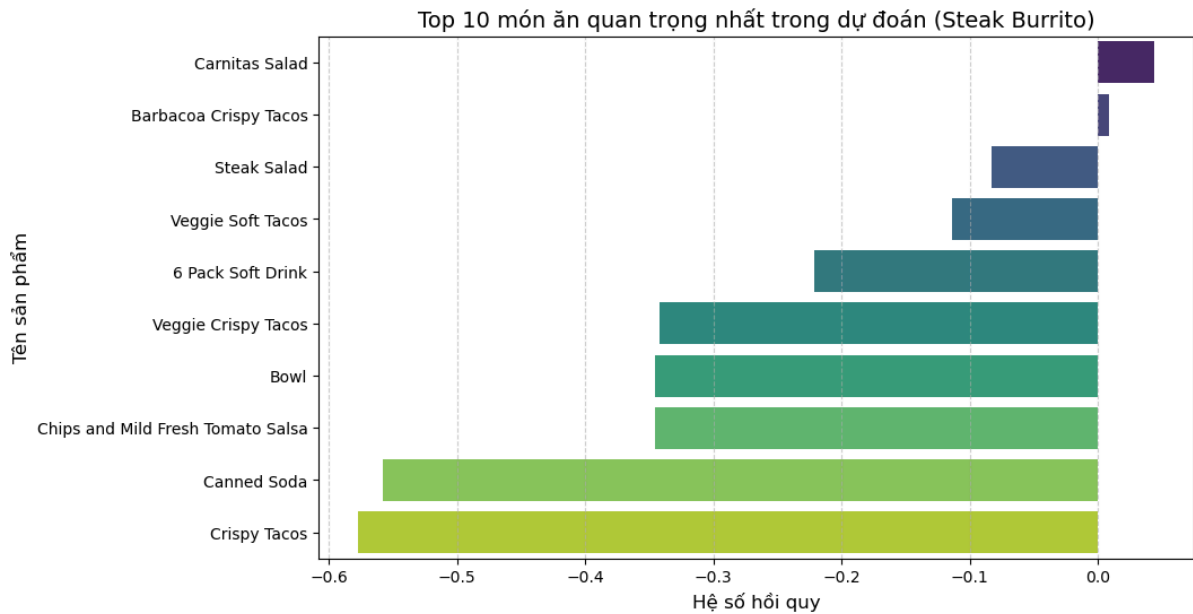
Món ăn có ảnh hưởng tiêu cực lớn nhất:

- **Crispy Tacos:** Ảnh hưởng tiêu cực lớn nhất, với hệ số khoảng -0.4 . Sự hiện diện của món này làm giảm khả năng khách hàng chọn Chips and Guacamole.
- Các món khác như **Chicken Soft Tacos**, **Veggie Bowl**, và **Bowl** cũng có ảnh hưởng tiêu cực đáng kể.

Cần nhắc thiết kế thực đơn:

- Các món ăn có hệ số dương (tích cực) nên được kết hợp quảng bá hoặc bán kèm với Chips and Guacamole để tối ưu doanh thu.
- Các món có hệ số âm nên được kiểm tra kỹ lưỡng để tránh làm giảm khả năng khách hàng chọn mục tiêu.

d. Steak Burrito



Món ăn có ảnh hưởng tích cực lớn nhất:

- Carnitas Salad: Có hệ số hồi quy dương cao nhất (~0.55). Điều này cho thấy khi món này xuất hiện, khả năng khách hàng chọn *Steak Burrito* tăng mạnh.
- Barbacoa Crispy Tacos và Steak Salad: Cũng có ảnh hưởng tích cực đáng kể, với hệ số lần lượt nhỏ hơn một chút.

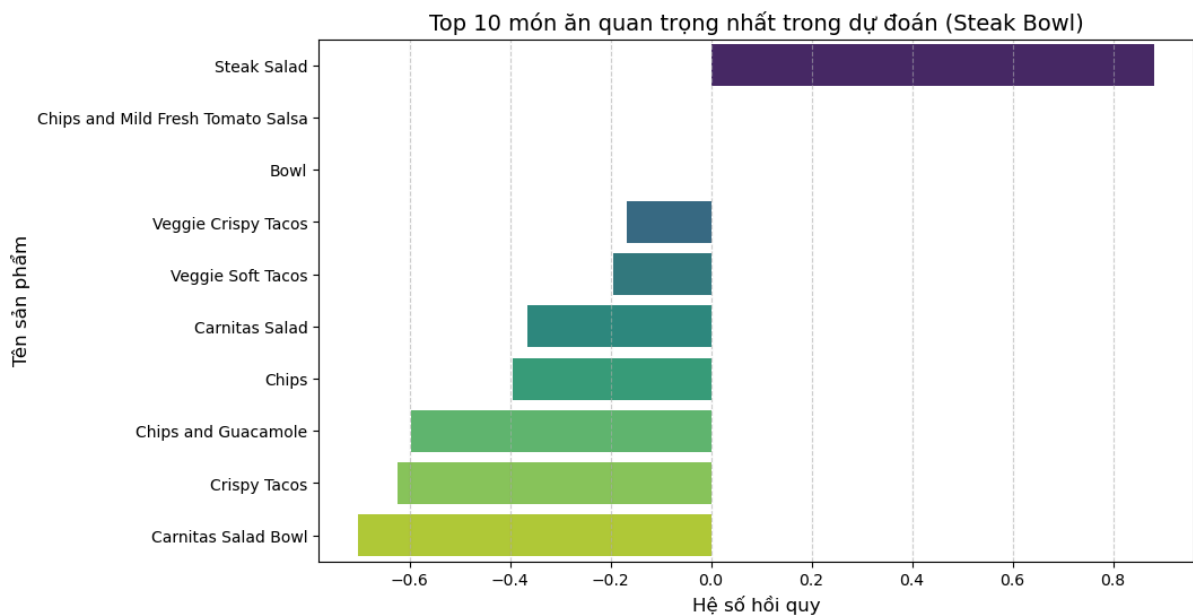
Món ăn có ảnh hưởng tiêu cực lớn nhất:

- Crispy Tacos: Hệ số hồi quy âm lớn nhất (~-0.6). Khi món này xuất hiện, khả năng khách hàng chọn *Steak Burrito* giảm đáng kể.
- Canned Soda: Ảnh hưởng tiêu cực mạnh, với hệ số ~-0.5.
- Các món khác như Chips and Mild Fresh Tomato Salsa cũng có ảnh hưởng tiêu cực đáng kể, nhưng mức độ thấp hơn.

3. Cân nhắc thiết kế thực đơn:

- Quảng bá món ăn tích cực:
 - Các món như Carnitas Salad, Barbacoa Crispy Tacos, và Steak Salad nên được kết hợp quảng bá hoặc bán kèm với *Steak Burrito* để tăng doanh số.
- Xem xét món ăn tiêu cực:
 - Các món như Crispy Tacos và Canned Soda có thể cần kiểm tra kỹ lưỡng. Chúng có thể không phù hợp để bán kèm với *Steak Burrito*, vì làm giảm khả năng khách hàng chọn món chính này.

e. Steak Bowl



Món ăn có ảnh hưởng tích cực lớn nhất:

- Steak Salad: Có hệ số hồi quy cao nhất (~0.8), nghĩa là nếu món này xuất hiện, khả năng khách hàng chọn Steak Bowl tăng mạnh.
- Crispy Tacos và Chips and Guacamole cũng có ảnh hưởng tích cực đáng kể, góp phần làm tăng xác suất dự đoán.

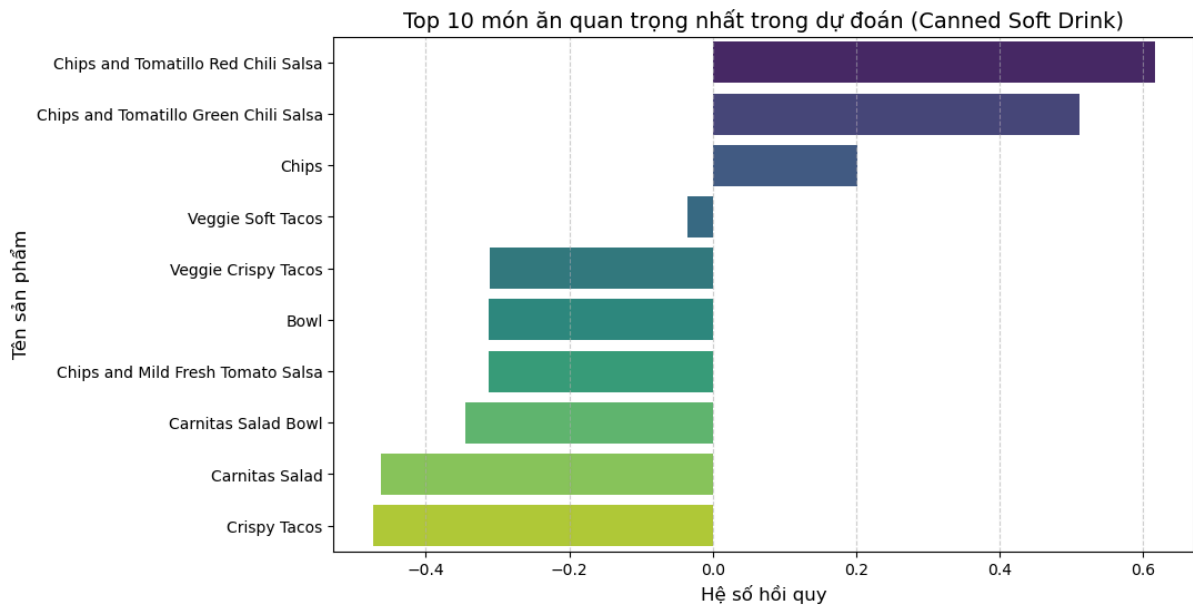
Món ăn có ảnh hưởng tiêu cực lớn nhất:

- Carnitas Salad Bowl: Ảnh hưởng tiêu cực lớn nhất, với hệ số khoảng -0.6. Sự hiện diện của món này làm giảm khả năng khách hàng chọn Steak Bowl.
- Các món khác như Chips và Veggie Crispy Tacos cũng có ảnh hưởng tiêu cực đáng kể, nhưng mức độ nhỏ hơn.

Cần nhắc thiết kế thực đơn:

- Các món có hệ số dương (tích cực) như Steak Salad, Crispy Tacos, và Chips and Guacamole nên được kết hợp quảng bá hoặc bán kèm với Steak Bowl để tối ưu doanh thu.
- Các món có hệ số âm (tiêu cực) như Carnitas Salad Bowl và Chips cần được kiểm tra kỹ để tránh làm giảm khả năng khách hàng chọn Steak Bowl.

f. Canned Soft Drink



Món ăn có ảnh hưởng tích cực lớn nhất:

- Chips and Tomatillo Red Chili Salsa: Có hệ số hồi quy cao nhất (~0.6), nghĩa là nếu món này xuất hiện, khả năng khách hàng chọn Canned Soft Drink tăng mạnh.
- Chips and Tomatillo Green Chili Salsa cũng có ảnh hưởng tích cực đáng kể, góp phần làm tăng xác suất dự đoán.

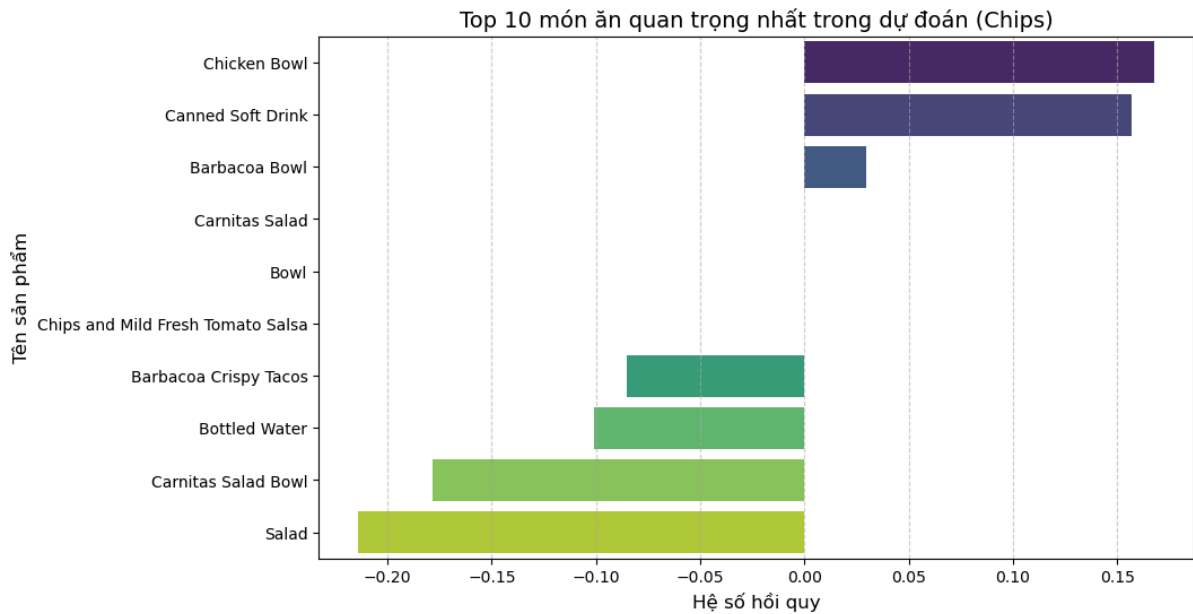
Món ăn có ảnh hưởng tiêu cực lớn nhất:

- Crispy Tacos: Ảnh hưởng tiêu cực lớn nhất, với hệ số khoảng -0.5. Sự hiện diện của món này làm giảm khả năng khách hàng chọn Canned Soft Drink.
- Các món khác như Carnitas Salad Bowl và Chips and Mild Fresh Tomato Salsa cũng có ảnh hưởng tiêu cực đáng kể, nhưng mức độ nhỏ hơn.

Cần nhắc thiết kế thực đơn:

- Quảng bá hoặc bán kèm: Các món có hệ số dương (tích cực) như Chips and Tomatillo Red Chili Salsa và Chips and Tomatillo Green Chili Salsa nên được kết hợp với Canned Soft Drink để tăng doanh thu.
- Kiểm tra kỹ lưỡng: Các món có hệ số âm (tiêu cực) như Crispy Tacos và Carnitas Salad Bowl cần được xem xét cẩn thận để tránh ảnh hưởng đến việc khách hàng chọn Canned Soft Drink.

g. Chips



Món ăn có ảnh hưởng tích cực lớn nhất:

- Chicken Bowl: Có hệ số hồi quy cao nhất (~ 0.6), nghĩa là khi món này xuất hiện, khả năng khách hàng chọn Chips tăng đáng kể.
- Canned Soft Drink cũng có ảnh hưởng tích cực đáng kể, góp phần tăng xác suất dự đoán.

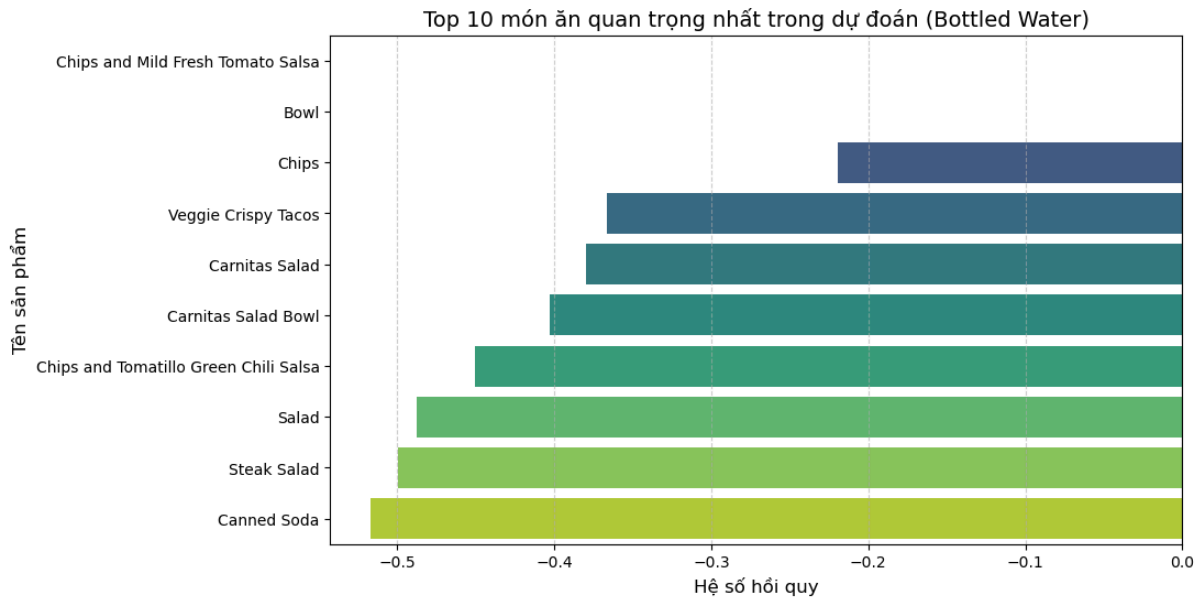
Món ăn có ảnh hưởng tiêu cực lớn nhất:

- Carnitas Salad Bowl: Ảnh hưởng tiêu cực lớn nhất (~ -0.3), sự hiện diện của món này làm giảm khả năng khách hàng chọn Chips.
- Các món khác như Bottled Water và Salad cũng có ảnh hưởng tiêu cực nhỏ hơn.

Cần nhắc thiết kế thực đơn:

- Quảng bá hoặc bán kèm: Các món có hệ số dương (tích cực) như Chicken Bowl và Canned Soft Drink nên được kết hợp với Chips để tối ưu hóa doanh thu.
- Kiểm tra kỹ lưỡng: Các món có hệ số âm (tiêu cực) như Carnitas Salad Bowl và Bottled Water cần được cân nhắc cẩn thận để tránh ảnh hưởng đến việc khách hàng chọn Chips.

h. Bottled Water



*Các món tăng khả năng dự đoán Bottled Water (hệ số dương lớn):

Không có món nào trong biểu đồ có hệ số dương, tất cả đều có ảnh hưởng tiêu cực đến khả năng dự đoán Bottled Water.

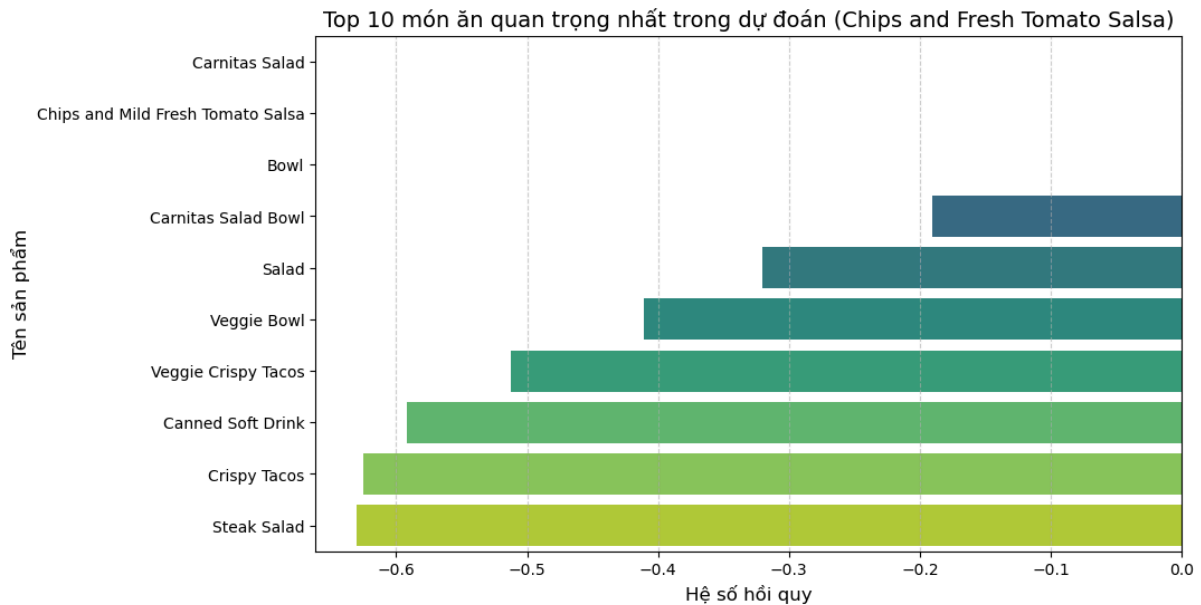
*Các món giảm khả năng dự đoán Bottled Water (hệ số âm lớn):

- Canned Soda: Ảnh hưởng tiêu cực mạnh nhất (~ -0.5), làm giảm đáng kể khả năng khách hàng chọn Bottled Water.
- Steak Salad và Salad: Cũng có ảnh hưởng tiêu cực đáng kể (~ -0.4), làm giảm xác suất chọn Bottled Water.
- Carnitas Salad Bowl và Chips: Góp phần giảm nhẹ hơn (~ -0.2 đến -0.3).

*Ý nghĩa thực tế:

- Món ảnh hưởng tiêu cực lớn như Canned Soda và Steak Salad cho thấy khách hàng chọn các món này ít có khả năng mua thêm Bottled Water.
- Món ảnh hưởng tiêu cực nhỏ như Chips và Veggie Crispy Tacos có thể được sử dụng làm món phụ, vì ảnh hưởng của chúng không quá lớn, giúp đa dạng hóa thực đơn.

i. Chips and Fresh Tomato Salsa



Các món tăng khả năng dự đoán Bottled Water (hệ số dương lớn):

- Không có món nào trong biểu đồ có hệ số dương. Tất cả các món được hiển thị đều có hệ số âm, nghĩa là chúng ảnh hưởng tiêu cực đến khả năng dự đoán Bottled Water.

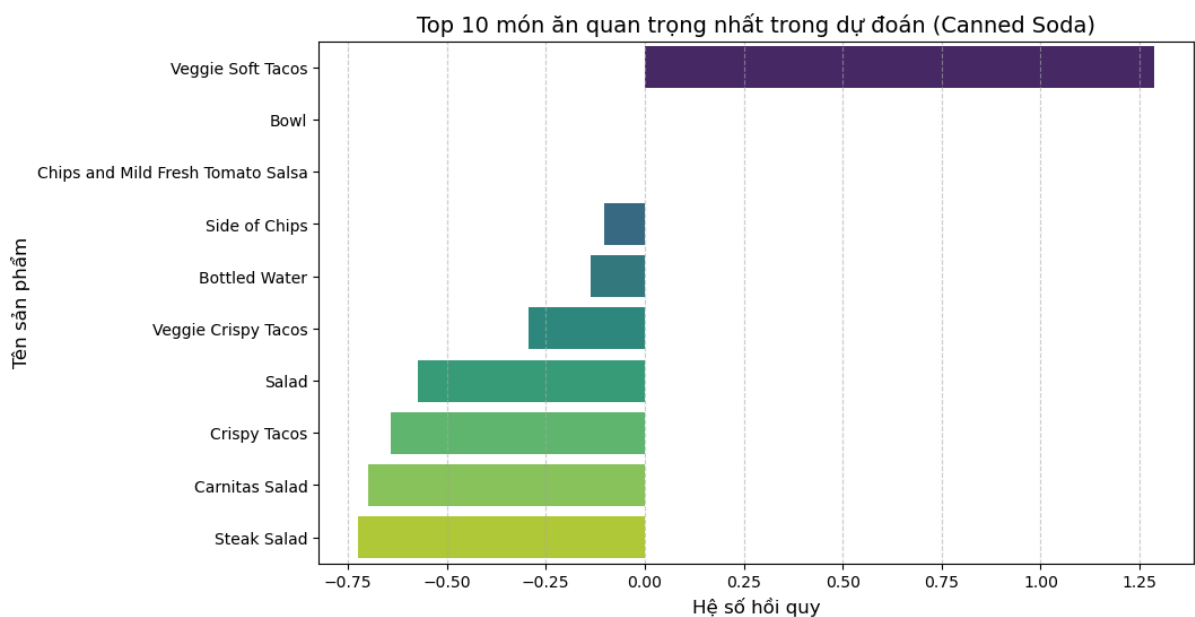
Các món giảm khả năng dự đoán Bottled Water (hệ số âm lớn):

- Canned Soda: Ảnh hưởng tiêu cực mạnh nhất (~ -0.6), làm giảm đáng kể khả năng khách hàng chọn Bottled Water.
- Steak Salad và Crispy Tacos: Có ảnh hưởng tiêu cực đáng kể (~ -0.4 đến -0.5), làm giảm xác suất chọn Bottled Water.
- Veggie Crispy Tacos và Canned Soft Drink: Góp phần giảm nhẹ hơn (~ -0.3 đến -0.4).
- Các món như Veggie Bowl, Chips, và Carnitas Salad Bowl: Ảnh hưởng tiêu cực ở mức trung bình (~ -0.2 đến -0.3).

Ý nghĩa thực tế:

- Các món ảnh hưởng tiêu cực lớn (Canned Soda, Steak Salad): Khách hàng chọn các món này ít có khả năng mua thêm Bottled Water. Do đó, nếu mục tiêu là tăng doanh số Bottled Water, cần giảm sự phụ thuộc hoặc quảng bá thêm Bottled Water cùng các món này.
- Các món ảnh hưởng tiêu cực nhỏ (Veggie Crispy Tacos, Chips): Những món này có thể được sử dụng làm món phụ vì ảnh hưởng của chúng không quá lớn. Điều này giúp đa dạng hóa thực đơn mà không giảm đáng kể xác suất bán Bottled Water.

j. Canned Soda



Món ăn có ảnh hưởng tích cực lớn nhất:

- Veggie Soft Tacos: Có hệ số hồi quy cao nhất (~1.2), nghĩa là sự xuất hiện của món này làm tăng mạnh khả năng khách hàng chọn Canned Soda.
- Bowl: Cũng có ảnh hưởng tích cực đáng kể (~0.6).

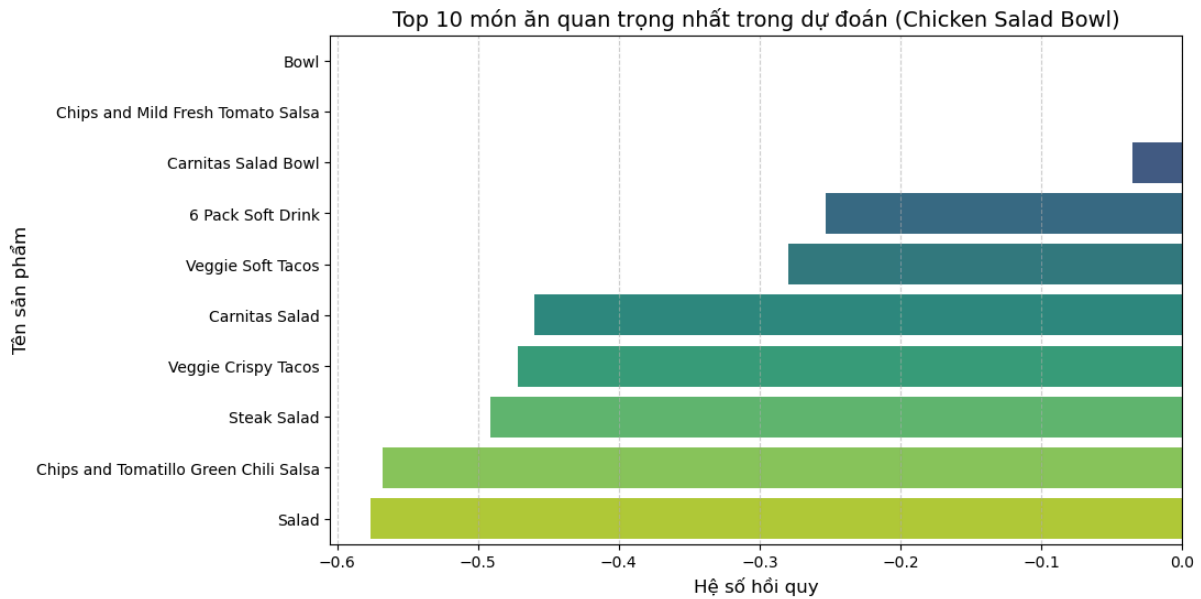
Món ăn có ảnh hưởng tiêu cực lớn nhất:

- Steak Salad: Ảnh hưởng tiêu cực mạnh nhất (~-0.8).
- Crispy Tacos và Salad: Cũng có ảnh hưởng tiêu cực đáng kể (~-0.6).

Cần nhắc thiết kế thực đơn:

- Quảng bá hoặc bán kèm: Các món như Veggie Soft Tacos và Bowl nên được kết hợp với Canned Soda để tăng doanh số.
- Kiểm tra kỹ lưỡng: Các món như Steak Salad và Crispy Tacos cần xem xét cách kết hợp để tránh giảm doanh thu từ Canned Soda.

k. Chicken Salad Bowl



Món ăn có ảnh hưởng tiêu cực lớn nhất:

1. Salad:
 - Hệ số hồi quy âm lớn nhất (~ -0.6), cho thấy món này làm giảm mạnh khả năng chọn Chicken Salad Bowl. Đây có thể là món cạnh tranh trực tiếp do tính chất tương tự.
2. Chips and Tomatillo Green Chili Salsa:
 - Hệ số âm đáng kể (~ -0.5), ám chỉ rằng món phụ này có thể là lựa chọn thay thế hoặc không kết hợp với Chicken Salad Bowl.
3. Steak Salad:
 - Hệ số âm (~ -0.4), tiếp tục củng cố vai trò là một món cạnh tranh trong cùng nhóm món ăn chính.

Món ăn có ảnh hưởng tiêu cực ít hơn:

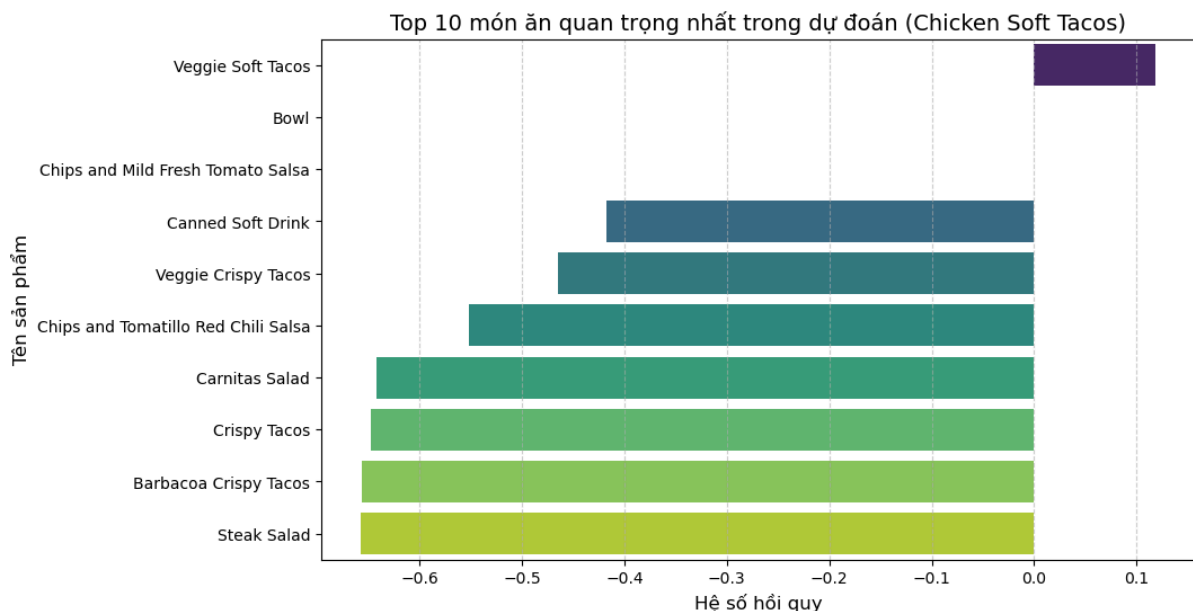
- Các món như Bowl và Chips and Mild Fresh Tomato Salsa có hệ số hồi quy âm nhỏ hơn (~ -0.1 đến ~ -0.2), cho thấy tác động không đáng kể đến khả năng chọn Chicken Salad Bowl.

Cần nhắc thiết kế thực đơn

- Quảng bá hoặc kết hợp món ăn bổ trợ: Các món có hệ số tiêu cực ít hơn như Bowl và Veggie Soft Tacos có thể được tận dụng để tạo combo với Chicken Salad Bowl, nhằm tăng doanh số mà không ảnh hưởng đến lựa chọn chính.
- Xử lý món ăn cạnh tranh:

- Salad và Steak Salad: Điều chỉnh giá hoặc thêm các thành phần đặc trưng để làm nổi bật sự khác biệt của Chicken Salad Bowl.
- Chips and Tomatillo Green Chili Salsa: Hạn chế khuyến mãi hoặc kết hợp món này với Chicken Salad Bowl nếu không hỗ trợ lẫn nhau.
- Chiến lược định vị món chính:
 - Làm nổi bật lợi thế của Chicken Salad Bowl (ví dụ: hương vị, thành phần đặc biệt, hoặc dinh dưỡng cao hơn).
 - Đẩy mạnh quảng bá trong các khung giờ phù hợp (ví dụ: giờ ăn trưa, khi khách hàng có xu hướng chọn salad).

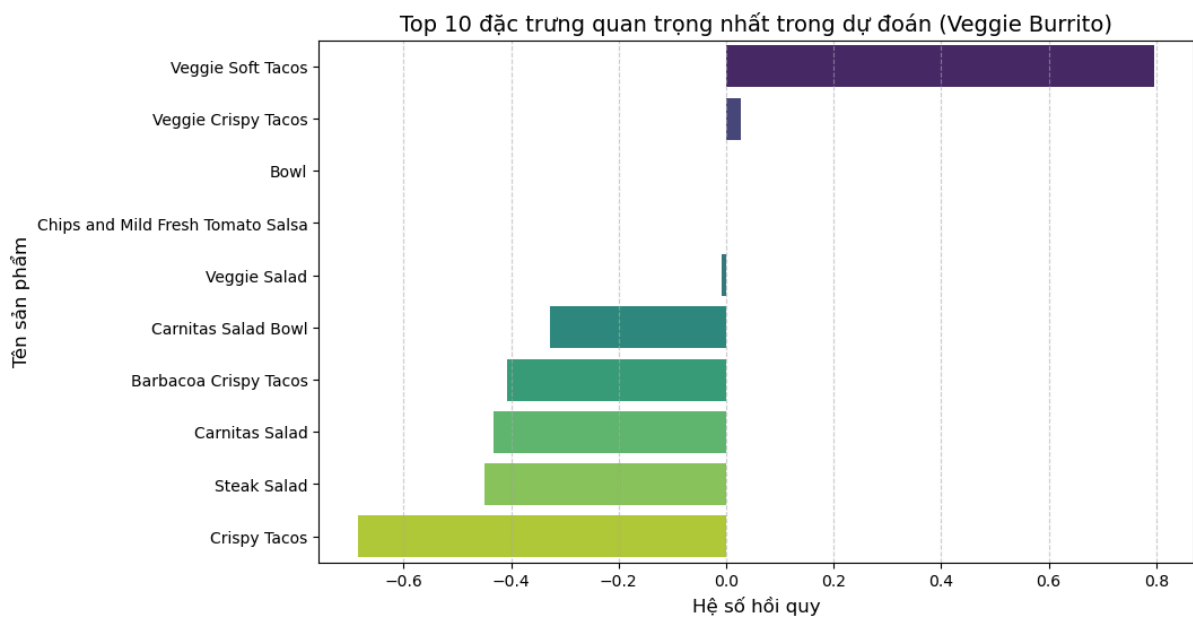
1. Chicken Soft Tacos



- Món ăn có ảnh hưởng tiêu cực lớn nhất: Món ăn "Steak Salad" có ảnh hưởng tiêu cực lớn nhất đến dự đoán (hệ số hồi quy thấp nhất).
- Món ăn có ảnh hưởng tích cực lớn nhất: Món ăn "Veggie Soft Tacos" có ảnh hưởng tích cực lớn nhất đến dự đoán (hệ số hồi quy cao nhất).
- Cân nhắc thiết kế thực đơn:
 - Nên ưu tiên các món như Veggie Soft Tacos trong chiến lược thực đơn vì chúng có ảnh hưởng tích cực đến dự đoán kết quả (có thể là mức độ phổ biến hoặc khả năng gợi ý tích cực).
 - Hạn chế hoặc điều chỉnh cách quảng bá các món như Steak Salad vì chúng có tác động tiêu cực đến kết quả dự đoán.

- Xem xét các món như Crispy Tacos và Barbacoa Crispy Tacos, vì chúng có ảnh hưởng gần mức trung bình, để tìm cơ hội cải thiện hiệu suất hoặc giảm thiểu tác động tiêu cực.

m. Veggie Burrito



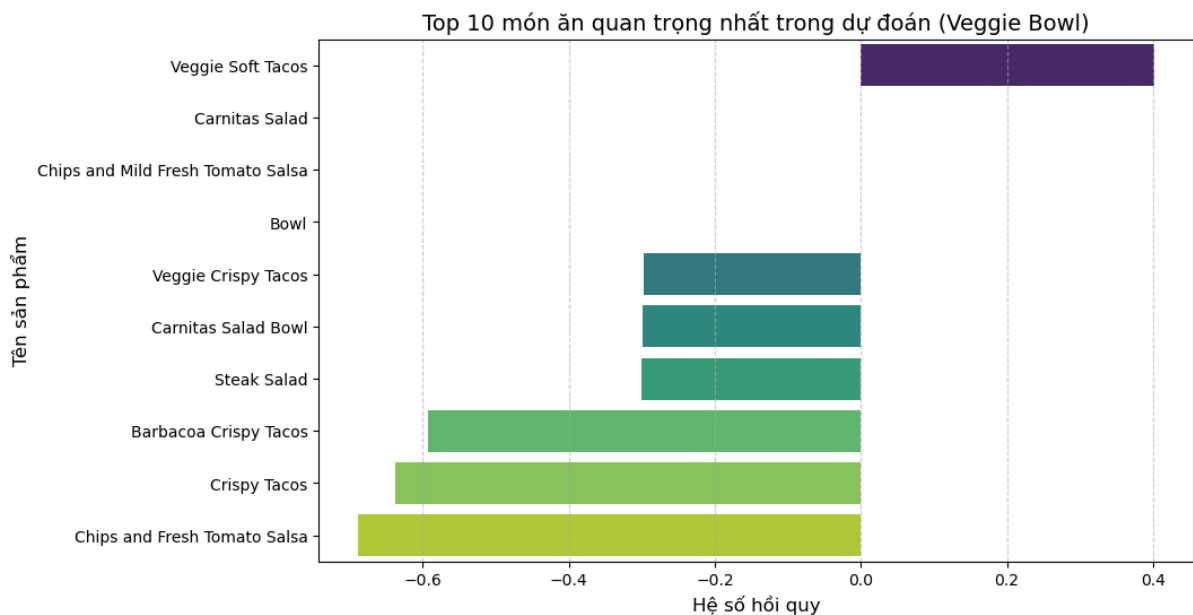
Món ăn ảnh hưởng tích cực:

- Veggie Soft Tacos và Veggie Crispy Tacos có tác động tích cực mạnh nhất đến dự đoán cho Veggie Burrito, với hệ số hồi quy cao.

Món ăn ảnh hưởng tiêu cực:

- Crispy Tacos và Steak Salad có tác động tiêu cực lớn nhất, làm giảm xác suất dự đoán cho Veggie Burrito.
- Cần nhắc thiết kế thực đơn:
 - Điều chỉnh món có ảnh hưởng tiêu cực: Xem xét cách cải thiện sự hấp dẫn của các món như Crispy Tacos và Steak Salad. Có thể điều chỉnh cách chế biến, thành phần, hoặc nhóm khách hàng mục tiêu để giảm tác động tiêu cực lên dự đoán.
 - Tập trung vào món liên quan: Kết hợp quảng bá các món như Veggie Soft Tacos và Veggie Burrito trong các combo hoặc thực đơn, vì chúng có ảnh hưởng tích cực lẫn nhau.

n. Veggie Bowl



Món ăn ảnh hưởng tích cực:

- Veggie Soft Tacos có ảnh hưởng tích cực lớn nhất đến dự đoán cho Veggie Bowl, với hệ số hồi quy cao nhất.
- Các món như Carnitas Salad và Chips and Mild Fresh Tomato Salsa cũng có tác động tích cực đáng kể.

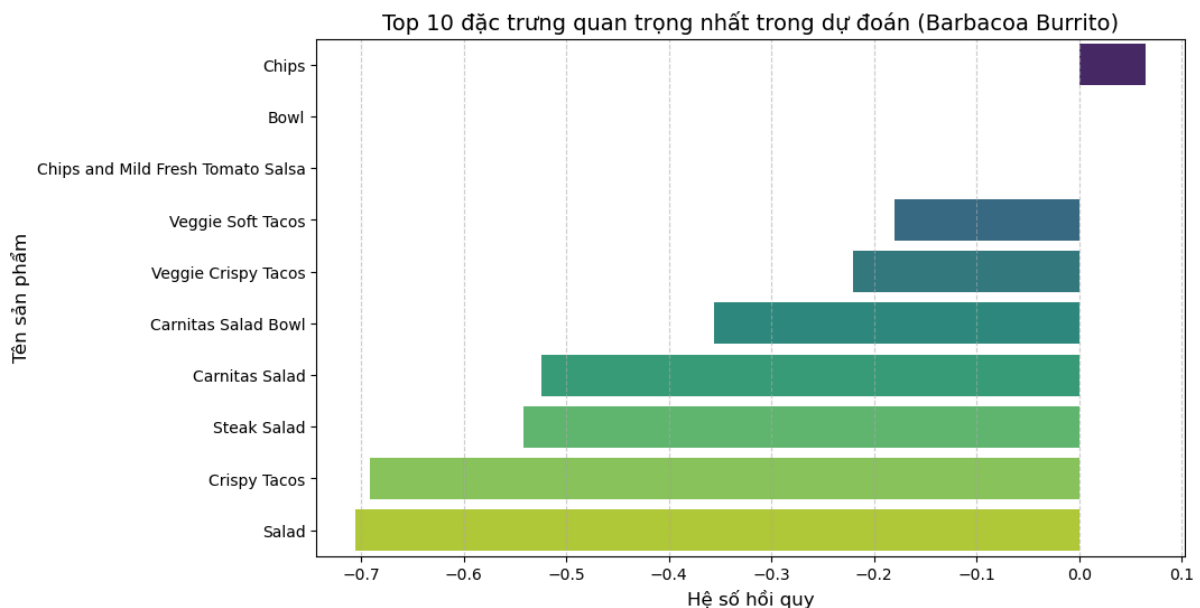
Món ăn ảnh hưởng tiêu cực:

- Chips and Fresh Tomato Salsa và Crispy Tacos có ảnh hưởng tiêu cực lớn nhất, làm giảm xác suất dự đoán cho Veggie Bowl.
- Các món như Barbacoa Crispy Tacos và Steak Salad cũng góp phần làm giảm khả năng dự đoán.

Cân nhắc thiết kế thực đơn:

- Tăng cường các món Veggie: Kết hợp các món như Veggie Soft Tacos và Veggie Bowl trong thực đơn để tăng sự liên kết giữa chúng, nhằm tới nhóm khách hàng ưa chuộng món ăn lành mạnh.
- Điều chỉnh món tiêu cực: Đánh giá lại các món có ảnh hưởng tiêu cực như Crispy Tacos và Chips and Fresh Tomato Salsa để giảm cạnh tranh trực tiếp với Veggie Bowl, hoặc thay đổi cách phối hợp các món này trong thực đơn.

o. Barbacoa Burrito



Món ăn ảnh hưởng tích cực:

- Chips and Mild Fresh Tomato Salsa có ảnh hưởng tích cực lớn nhất đối với dự đoán Barbacoa Burrito, thể hiện sự liên kết mạnh mẽ giữa món này và Burrito.
- Veggie Soft Tacos và Veggie Crispy Tacos cũng đóng vai trò quan trọng, nhưng mức độ tác động tích cực ít hơn.

Món ăn ảnh hưởng tiêu cực:

- Crispy Tacos và Steak Salad có ảnh hưởng tiêu cực lớn nhất, làm giảm xác suất dự đoán Barbacoa Burrito.
- Carnitas Salad Bowl và Carnitas Salad cũng có tác động tiêu cực đáng kể, cần lưu ý khi thiết kế thực đơn.

Cần nhắc thiết kế thực đơn:

- Tăng cường món tích cực: Khuyến khích khách hàng kết hợp Chips and Mild Fresh Tomato Salsa với Barbacoa Burrito trong combo hoặc thực đơn để tăng độ phổ biến và khả năng dự đoán chính xác.
- Điều chỉnh món tiêu cực: Xem xét thay đổi công thức hoặc cách trình bày các món như Crispy Tacos và Steak Salad, nhằm giảm tác động tiêu cực đến món Barbacoa Burrito.

- Tạo sự bổ sung: Tích hợp các món có tác động trung lập hoặc hơi tiêu cực (ví dụ: Veggie Tacos) vào thực đơn với món Barbacoa Burrito để tối ưu hóa sự kết hợp.