**TỐI ƯU HÓA DỊCH VỤ NHÀ HÀNG THÔNG QUA CÁC CÔNG CỤ TRÍ TUỆ KINH DOANH: MỘT NGHIÊN CỨU TRƯỜNG HỢP VỀ CHUỖI NHÀ HÀNG PIZZA B&P**

**TÓM TẮT**

Ngành công nghiệp nhà hàng đang phải đối mặt với những thách thức ngày càng tăng trong việc cải thiện hiệu quả hoạt động và sự hài lòng của khách hàng. Nghiên cứu này khám phá cách các công cụ Business Intelligence (BI) có thể tối ưu hóa các dịch vụ nhà hàng, tập trung vào chuỗi nhà hàng Pizza B&P như một nghiên cứu điển hình. Các mục tiêu chính là tăng cường tối ưu hóa thực đơn, tăng doanh số và cải thiện trải nghiệm của khách hàng bằng cách phân tích dữ liệu bán hàng, phản hồi của khách hàng và hiệu suất của nhân viên trong sáu tháng.

Các tập dữ liệu chính bao gồm hồ sơ bán hàng, thông tin nhân khẩu học của khách hàng, xếp hạng phản hồi và nhật ký hoạt động của nhân viên. Sử dụng các công cụ như Python, Tableau và Flask/Django, dự án áp dụng phân tích dữ liệu khám phá (EDA), mô hình dự đoán (ví dụ: hồi quy tuyến tính để dự báo doanh số) và phân khúc khách hàng thông qua cụm. Ngoài ra, các kỹ thuật xử lý ngôn ngữ tự nhiên (NLP) được sử dụng để phân tích phản hồi của khách hàng.

Nghiên cứu cung cấp thông tin chi tiết có thể hành động thông qua các báo cáo phân tích chi tiết và bảng điều khiển BI tương tác. Bảng điều khiển cung cấp hình ảnh trực quan theo thời gian thực về xu hướng bán hàng, phân khúc khách hàng và hiệu suất của nhân viên. Các mô hình dự báo xác định các món ăn bán chạy nhất và nhóm khách hàng, cho phép đưa ra các khuyến nghị chiến lược như chương trình khuyến mãi có mục tiêu và tối ưu hóa thời gian phục vụ. Một ứng dụng web tích hợp các tính năng này, trao quyền cho các nhà quản lý nhà hàng đưa ra quyết định dựa trên dữ liệu và nâng cao hiệu quả hoạt động.

Bằng cách triển khai các chiến lược do BI thúc đẩy này, nghiên cứu chứng minh tiềm năng tận dụng phân tích dữ liệu để cải thiện hiệu suất nhà hàng và sự hài lòng của khách hàng trong ngành dịch vụ thực phẩm cạnh tranh.

**CHƯƠNG 1: GIỚI THIỆU VỀ ĐỀ TÀI**

**1.1. Đề tài**

Tối ưu hóa dịch vụ nhà hàng thông qua Business Intelligence: Nghiên cứu tình huống tại chuỗi nhà hàng Pizza B&P.

**1.2. Mục tiêu**

Tối ưu hóa hoạt động kinh doanh của chuỗi nhà hàng Pizza B&P thông qua các công cụ BI bằng cách:

* **Dự đoán doanh số bán hàng** theo món ăn.
  + **Ý nghĩa:**
    - Hiểu được xu hướng tiêu thụ của từng loại pizza trong 30 ngày/1 năm tiếp theo.
    - Hỗ trợ quản lý nguồn cung nguyên liệu, đặc biệt là với các món bán chạy.
  + **Ứng dụng thực tế:**
    - Điều chỉnh số lượng nguyên liệu nhập vào để tránh lãng phí và thiếu hụt.
    - Lên kế hoạch khuyến mãi cho các loại pizza bán chậm để thúc đẩy doanh số.
* **Tối ưu hóa thực đơn** dựa trên phân tích dữ liệu bán hàng và phản hồi của khách hàng.
  + **Ý nghĩa:**
    - Phân tích dữ liệu dự báo để xác định các món bán chạy và bán chậm.
    - Sử dụng thông tin để cải thiện menu, tập trung vào các món được yêu thích.
  + **Ứng dụng thực tế:**
    - Loại bỏ hoặc điều chỉnh giá các món bán chậm.
    - Tăng sự đa dạng của các món bán chạy hoặc sáng tạo món mới dựa trên xu hướng.
* **Cải thiện hiệu suất dịch vụ**, giảm thời gian phục vụ.
  + **Ý nghĩa:**
    - Dự báo số lượng đơn hàng hàng ngày giúp tối ưu hóa lịch làm việc của nhân viên.
    - Giảm thời gian phục vụ thông qua việc chuẩn bị trước các món phổ biến.
  + **Ứng dụng thực tế:**
    - Tăng cường nhân sự vào các ngày dự báo có nhu cầu cao.
    - Sắp xếp nguyên liệu hợp lý để giảm thời gian chuẩn bị món.

**1.3. Lý do chọn đề tài**

Nhiều chuỗi nhà hàng gặp khó khăn trong việc:

* **Dự đoán nhu cầu khách hàng** và tối ưu hóa nguồn lực.
* **Hiểu rõ phản hồi khách hàng** để cải thiện dịch vụ.
* **Tăng doanh số** của các món ăn có hiệu suất kém.  
  ⟹ Nhóm chúng tôi quyết định sử dụng dữ liệu thực tế từ chuỗi nhà hàng Pizza B&P để trực tiếp hỗ trợ các mục tiêu trên, đồng thời cung cấp thông tin phong phú để triển khai các giải pháp BI.

**1.4. Phạm vi nghiên cứu**

Phân tích dữ liệu từ chuỗi nhà hàng Pizza B&P trong vòng **12 tháng**, bao gồm:

* Dữ liệu doanh số theo từng món ăn.
* Phản hồi của khách hàng.

**1.5. Kết quả dự kiến**

1. Báo cáo chi tiết với các phân tích và trực quan hóa dữ liệu.
2. Dashboard BI hiển thị doanh số và phản hồi của khách hàng.
3. Mô hình dự đoán doanh số và phân khúc khách hàng.
4. Đề xuất chiến lược cải thiện hiệu suất hoạt động của nhà hàng.

**CHƯƠNG 2: GIỚI THIỆU VỀ DỮ LIỆU**

**2.1. Mô tả dữ liệu**

Tệp **Expanded\_Pizza\_Sales.xlsx** bao gồm 48,620 dòng và 22 cột, cụ thể như sau:

1. **Thông tin đơn hàng**:
   * order\_details\_id, order\_id, pizza\_id, quantity.
2. **Thời gian đơn hàng**:
   * order\_date, order\_time.
3. **Thông tin giá và loại pizza**:
   * unit\_price, total\_price, pizza\_size, pizza\_category.
4. **Thành phần và tên món ăn**:
   * pizza\_ingredients, pizza\_name.
5. **Thông tin khách hàng**:
   * customer\_name, customer\_gender, customer\_dob.
6. **Phản hồi khách hàng**:
   * customer\_feedback, feedback\_platform, feedback\_date.
7. **Hoạt động nhân viên**:
   * employee\_id, employee\_activity.

**2.2. Mối quan hệ giữa các đặc trưng dữ liệu**

1. **Dữ liệu đơn hàng**:
   * order\_date: Từ ngày 01/01/2015 đến 31/12/2015.
   * order\_time: Có 16,382 giá trị khác nhau, ghi nhận rất chi tiết về thời gian.
   * quantity (trung bình): ~1.02, hầu hết khách hàng đặt 1 pizza/lần.
2. **Thông tin giá**:
   * unit\_price: Từ $9.75 đến $35.95, trung bình $16.49.
   * total\_price: Giá trị lớn nhất là $83, có thể do đơn hàng lớn.
3. **Chi tiết pizza**:
   * Có 32 loại pizza, chia thành 4 danh mục chính và 5 kích thước khác nhau.
   * Món bán chạy nhất: The Classic Deluxe Pizza.
4. **Phản hồi khách hàng**:
   * Có tổng cộng 64 khách hàng với các phản hồi đa dạng, ví dụ: “Rất thích không gian quán.”
   * customer\_gender: Gồm 3 nhóm giá trị: Nam, Nữ và nhóm “khác” hoặc không xác định.
5. **Hoạt động nhân viên**:
   * Bao gồm các nhiệm vụ như: “Đặt hàng”, “Giao hàng”, “Chuẩn bị món ăn” và “Chăm sóc khách hàng”.

**2.3. Lý do chọn dataset này thay vì dataset khác**

1. **Phù hợp với mục tiêu nghiên cứu**
   * Dataset này cung cấp dữ liệu toàn diện, liên quan trực tiếp đến mục tiêu nghiên cứu:
     + **Dự đoán doanh số**: Thông qua các trường dữ liệu như order\_date, order\_time, unit\_price, total\_price, và quantity.
     + **Phân tích hành vi khách hàng**: Với các trường customer\_feedback, customer\_gender, customer\_dob, và pizza\_preferences.
     + **Tối ưu hóa thực đơn**: Dựa vào thông tin doanh số (quantity) và thành phần món ăn (pizza\_ingredients).
2. **Mức độ chi tiết và tính thực tế cao**
   * Dataset ghi nhận **48,620 giao dịch thực tế** trong vòng 1 năm, đảm bảo:
     + **Tính toàn diện**: Bao quát mọi thời điểm trong năm (mùa cao điểm và thấp điểm).
     + **Phân tích xu hướng**: Xác định biến động doanh thu theo thời gian (ngày/giờ) và theo chi nhánh.
   * Dữ liệu phản ánh hoạt động kinh doanh thực tế, không phải dữ liệu giả lập.
3. **Đa dạng thông tin**
   * Dataset chứa không chỉ dữ liệu doanh số mà còn bao gồm:
     + **Thông tin khách hàng**: Phân tích nhóm khách hàng mục tiêu dựa trên độ tuổi, giới tính, hành vi.
     + **Phản hồi khách hàng**: Đánh giá sự hài lòng và cải thiện dịch vụ.
     + **Hoạt động nhân viên**: Đánh giá hiệu suất làm việc và tối ưu hóa thời gian phục vụ.
4. **Khả năng khai thác bằng các kỹ thuật BI**
   * Dataset phù hợp để áp dụng các kỹ thuật phân tích tiên tiến:
     + **Trực quan hóa dữ liệu**: Tableau có thể hiển thị xu hướng bán hàng, nhóm khách hàng, và hiệu suất dịch vụ.
     + **Phân tích nâng cao**:  
       ♣ **Dự báo doanh số**: Sử dụng các mô hình hồi quy.  
       ♣ **Phân cụm khách hàng**: Phân loại khách hàng theo hành vi.  
       ♣ **Phân tích cảm xúc**: Xác định mức độ hài lòng dựa trên phản hồi.

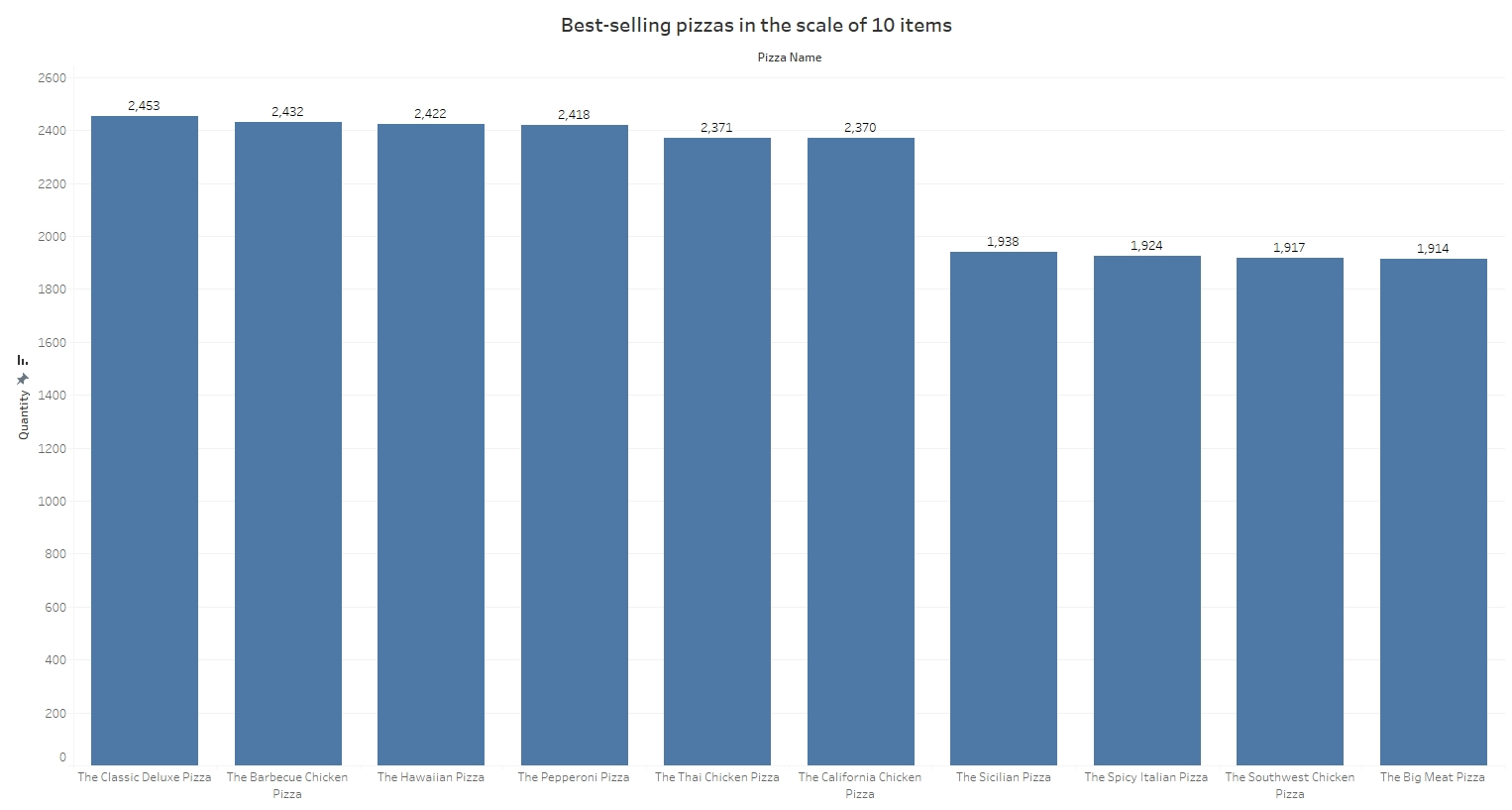
**CHƯƠNG 3: TIỀN XỬ LÝ VÀ TRỰC QUAN HÓA DỮ LIỆU**

**3.1. Tiền xử lý**

Các bước tiền xử lý dữ liệu:

1. **Kiểm tra và xử lý dữ liệu thiếu**
   * Xác định các cột có giá trị thiếu và quyết định:
     + Xóa cột (nếu thiếu quá nhiều).
     + Điền giá trị thay thế (trung bình, trung vị, hoặc mode).
2. **Xử lý dữ liệu không hợp lệ**
   * Kiểm tra các giá trị bất thường, ví dụ:
     + Giá trị âm trong các cột doanh thu hoặc số lượng.
     + Ngày tháng không hợp lệ hoặc nằm ngoài phạm vi mong muốn.
3. **Chuẩn hóa định dạng**
   * Chuyển các cột ngày tháng (order\_date và feedback\_date) sang định dạng chuẩn datetime.
   * Chuẩn hóa dữ liệu danh mục, ví dụ: phân loại kích thước pizza (pizza\_size) hoặc danh mục món ăn (pizza\_category).
4. **Tạo thêm các cột giá trị**
   * **Doanh thu bình quân trên mỗi pizza**: revenue\_per\_pizza = total\_price / quantity.
   * **Thời gian trong ngày**: Tạo cột order\_hour để phân tích giờ đặt hàng.

**3.2. Hình ảnh minh họa**



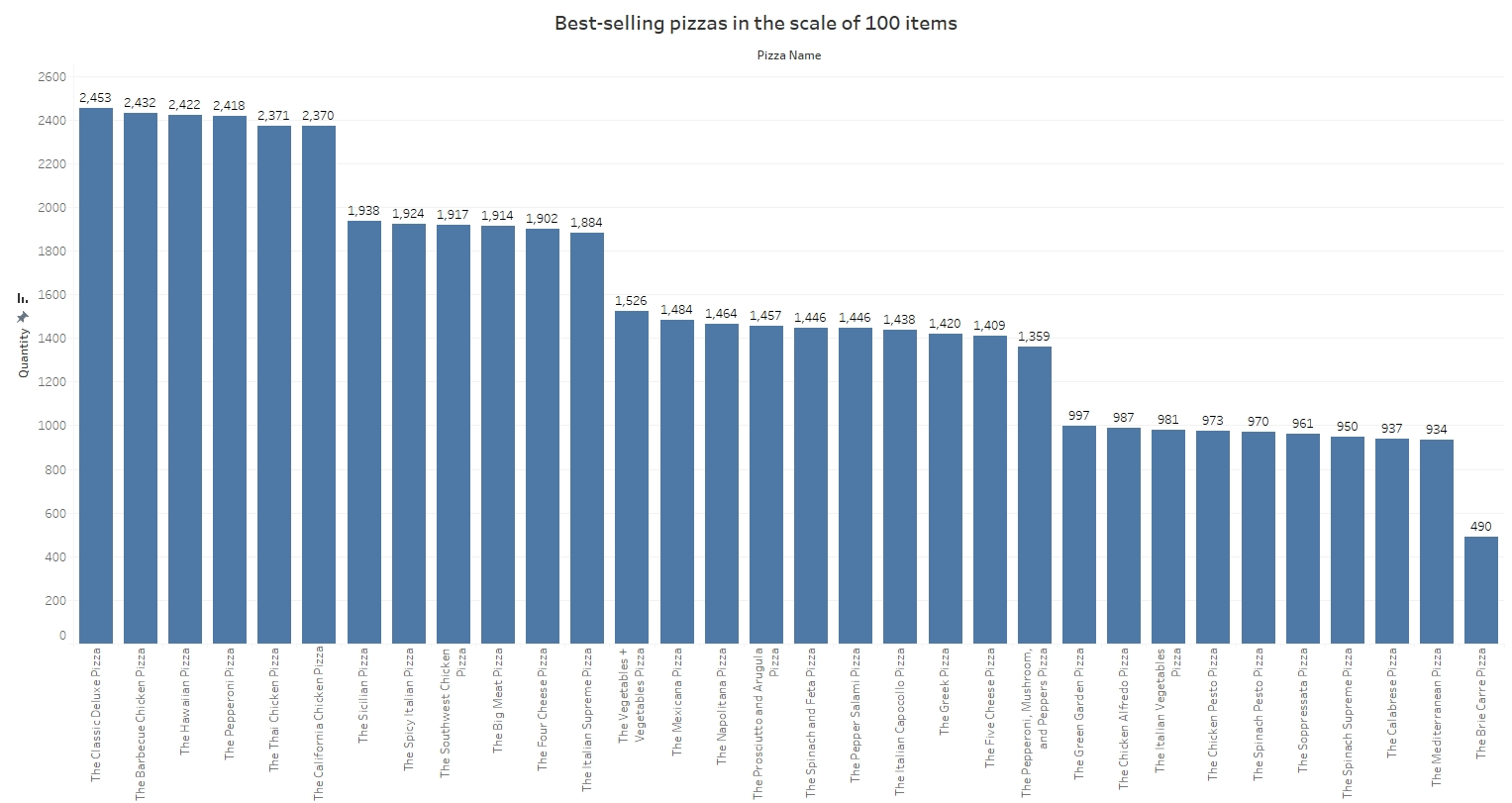
**Hình 3.2.1:** Biểu đồ các loại pizza bán chạy nhất theo quy mô 10 món

+ Thông tin: Biểu đồ này hiển thị 10 loại pizza bán chạy nhất theo số lượng

+ Ý nghĩa:

- Giúp nhà hàng hiểu được loại pizza nào được mọi người quan tâm. Do đó, nhà hàng biết rằng họ nên duy trì chất lượng của pizza.

Hơn nữa, nhà hàng có thể quyết định quảng bá pizza bằng cách giảm giá pizza vào những ngày đặc biệt.



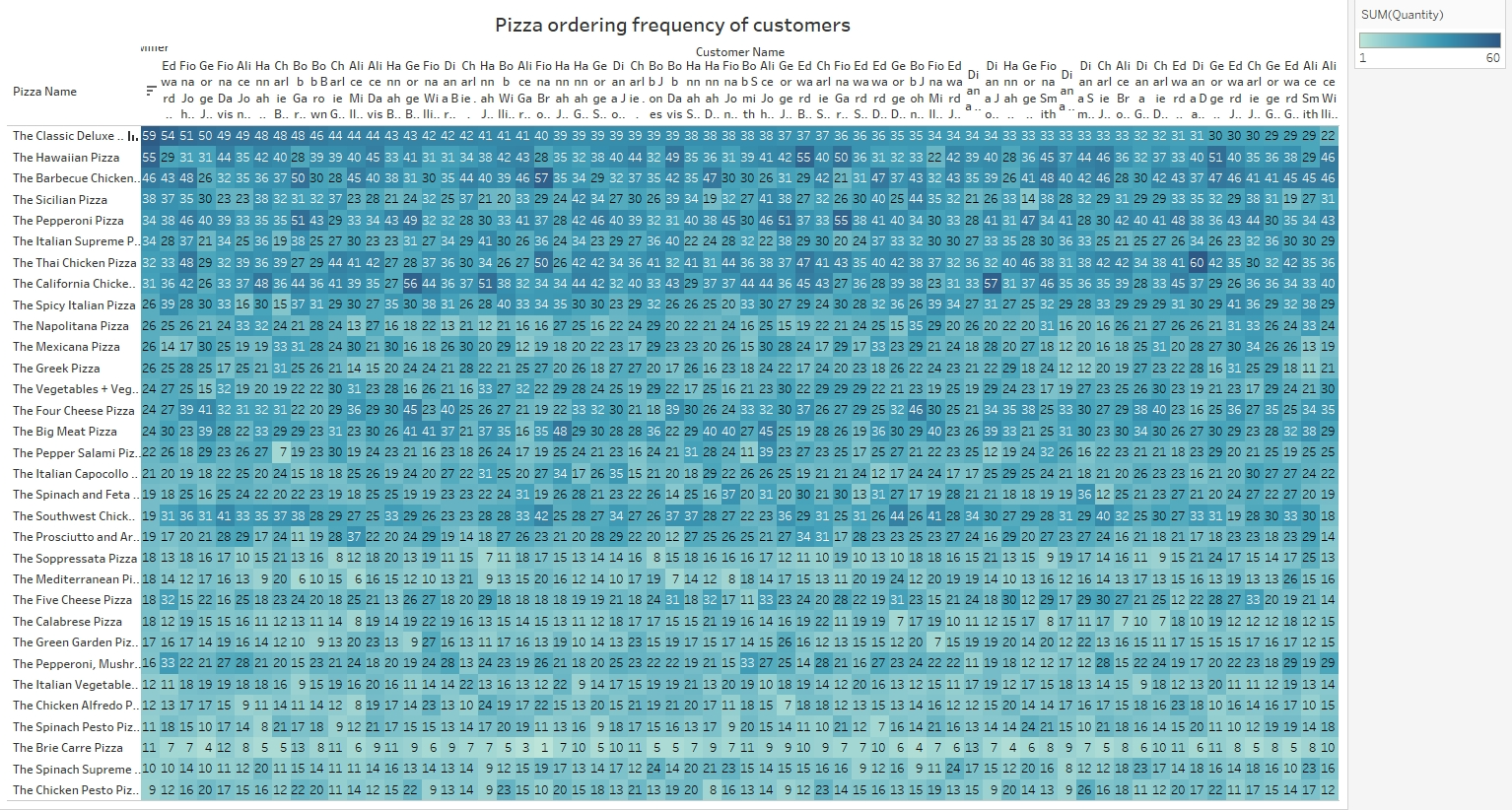
**Hình 3.2.2:** Biểu đồ các loại pizza bán chạy nhất theo quy mô 100 món

+ Thông tin: Biểu đồ này hiển thị 100 loại pizza bán chạy nhất theo số lượng

+ Ý nghĩa:

- Giúp nhà hàng hiểu được loại pizza nào được mọi người quan tâm. Do đó, nhà hàng biết rằng họ nên duy trì chất lượng của pizza có doanh số bán cao và cải thiện chất lượng của pizza có doanh số bán thấp.

- Với pizza có doanh số bán thấp, chủ nhà hàng hiểu rằng họ cần phải hành động, có thể là bắt đầu chương trình khuyến mãi để cải thiện doanh số bán và doanh thu của loại pizza đó hoặc thay đổi giá pizza để thu hút khách hàng đặt hàng.



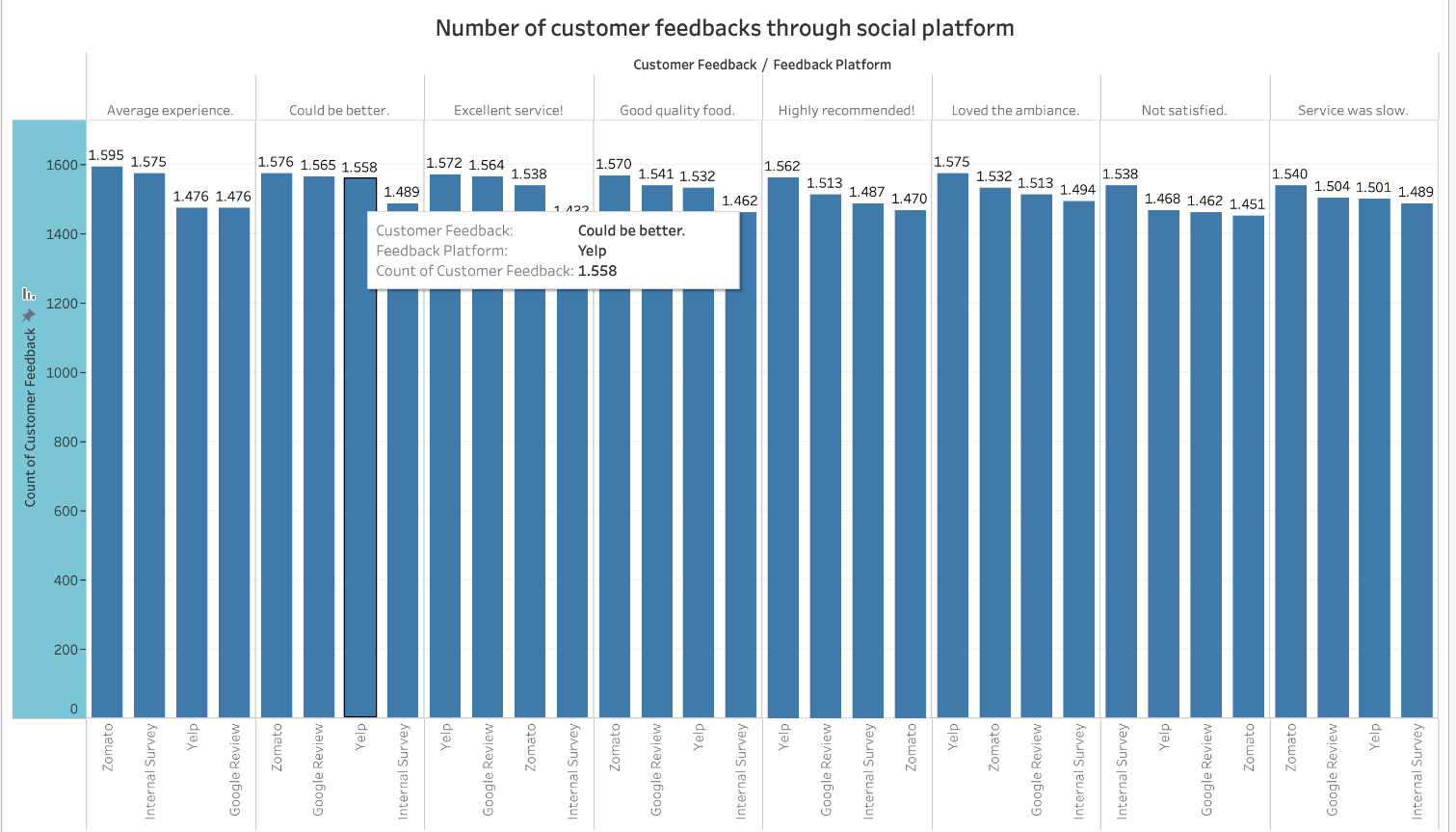
**Hình 3.2.3:** Tần suất đặt pizza của khách hàng

+ Thông tin: Con số này biểu thị tần suất đặt pizza của khách hàng đối với tất cả các loại pizza trong cửa hàng.

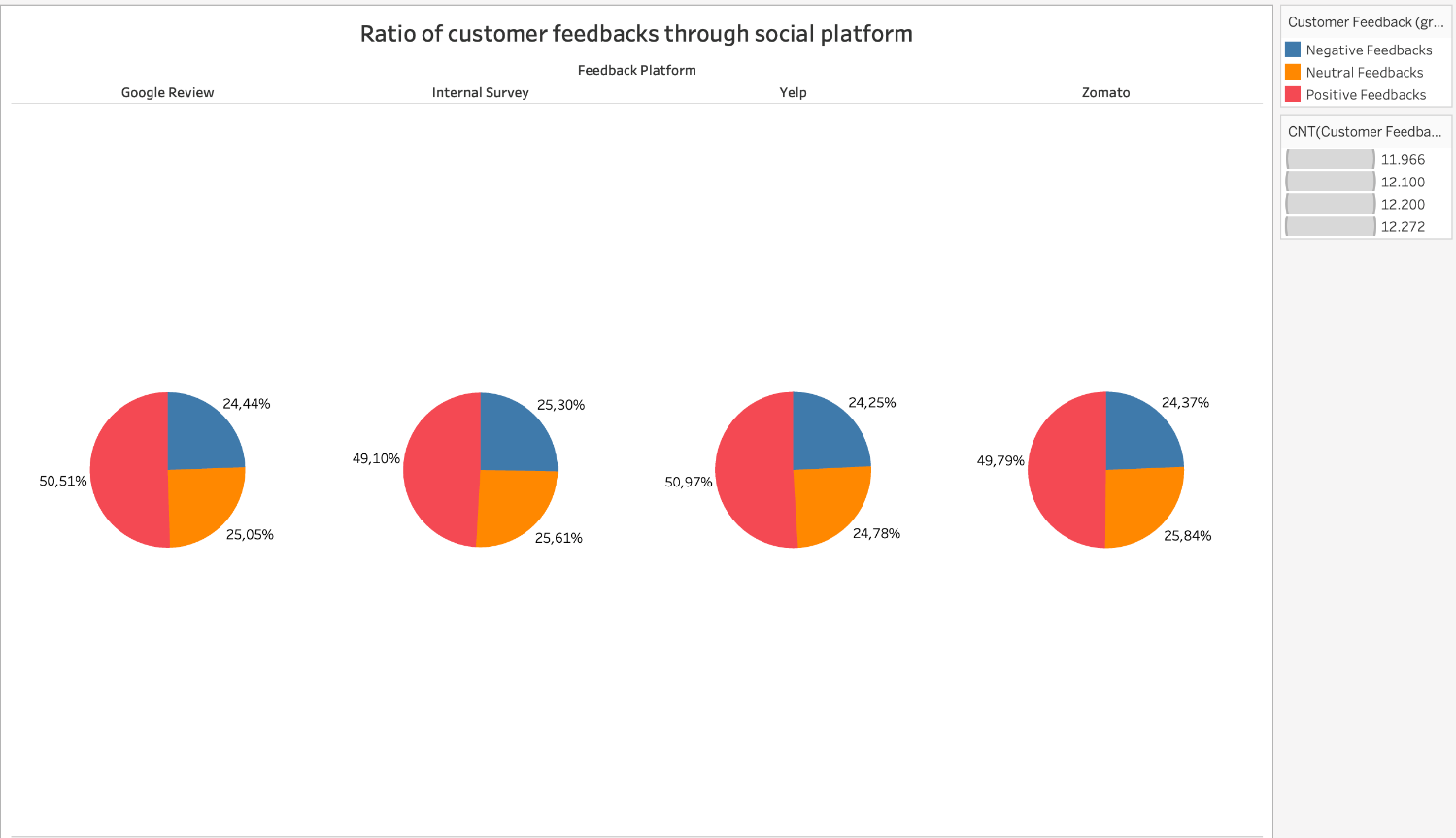
+ Ý nghĩa:

- Qua con số này, chủ nhà hàng có thể hiểu được tần suất gọi từng chiếc pizza của tất cả các khách hàng để hiểu khách hàng hơn.

- Do đó, nhà hàng có thể chuẩn bị một số phiếu mua hàng hoặc ưu đãi đặc biệt cho từng khách hàng với chiếc pizza mà họ gọi nhiều nhất.

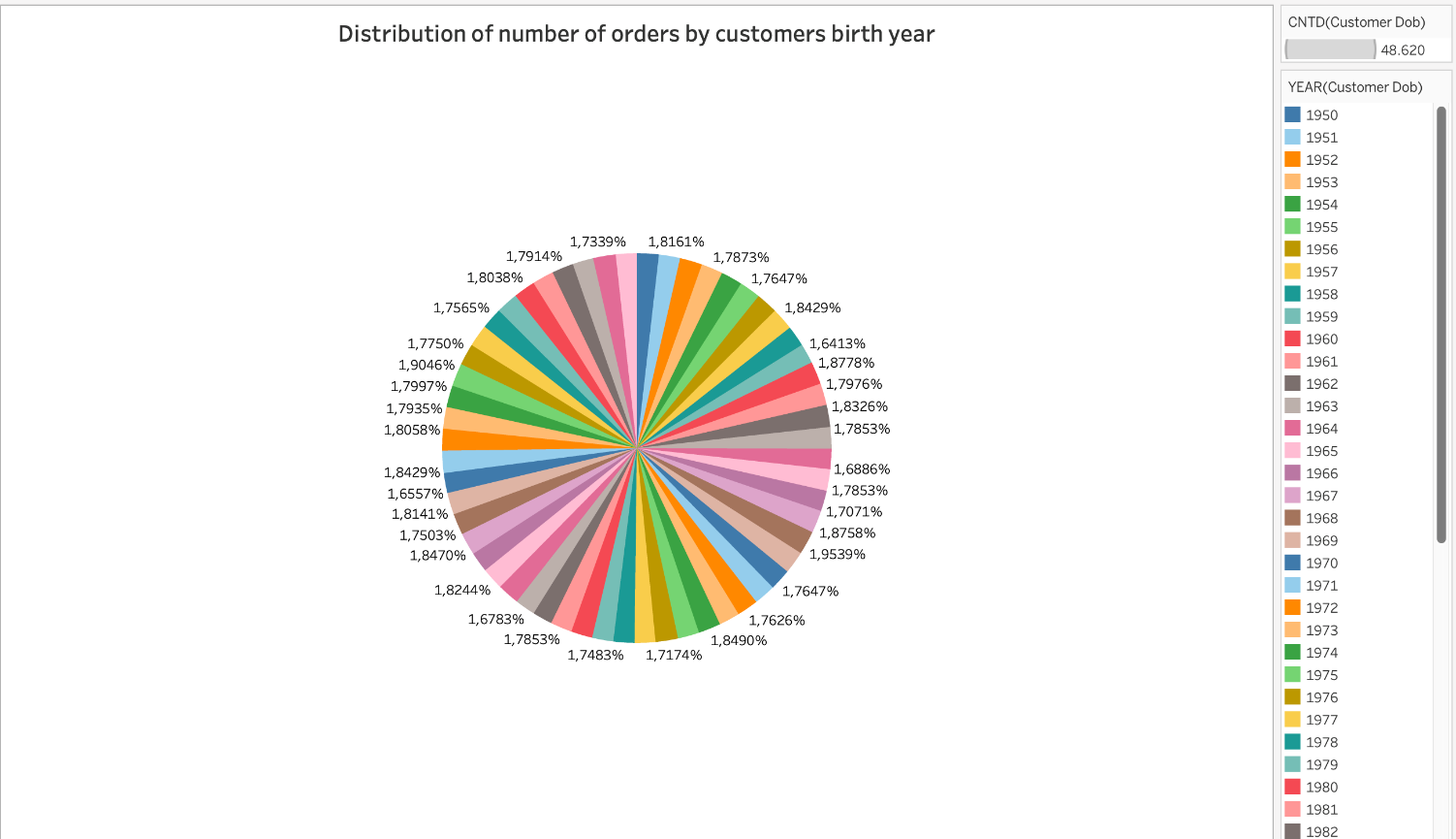
  
**Hình 3.2.4:** Biểu đồ cột thể hiện số lượng phản hồi qua các nền tảng xã hội

* **Thông tin:** Biểu đồ cột này thể hiện số lượng phản hồi qua bốn nền tảng xã hội bao gồm: Zomato, Yelp, Khảo sát nội bộ, Google Review.
* **Ý nghĩa:** Biểu đồ này giúp chủ nhà hàng hiểu được cảm nhận của khách hàng về nhà hàng. Do đó, chủ nhà hàng có thể hiểu được những khía cạnh mà nhà hàng cần cải thiện và duy trì.



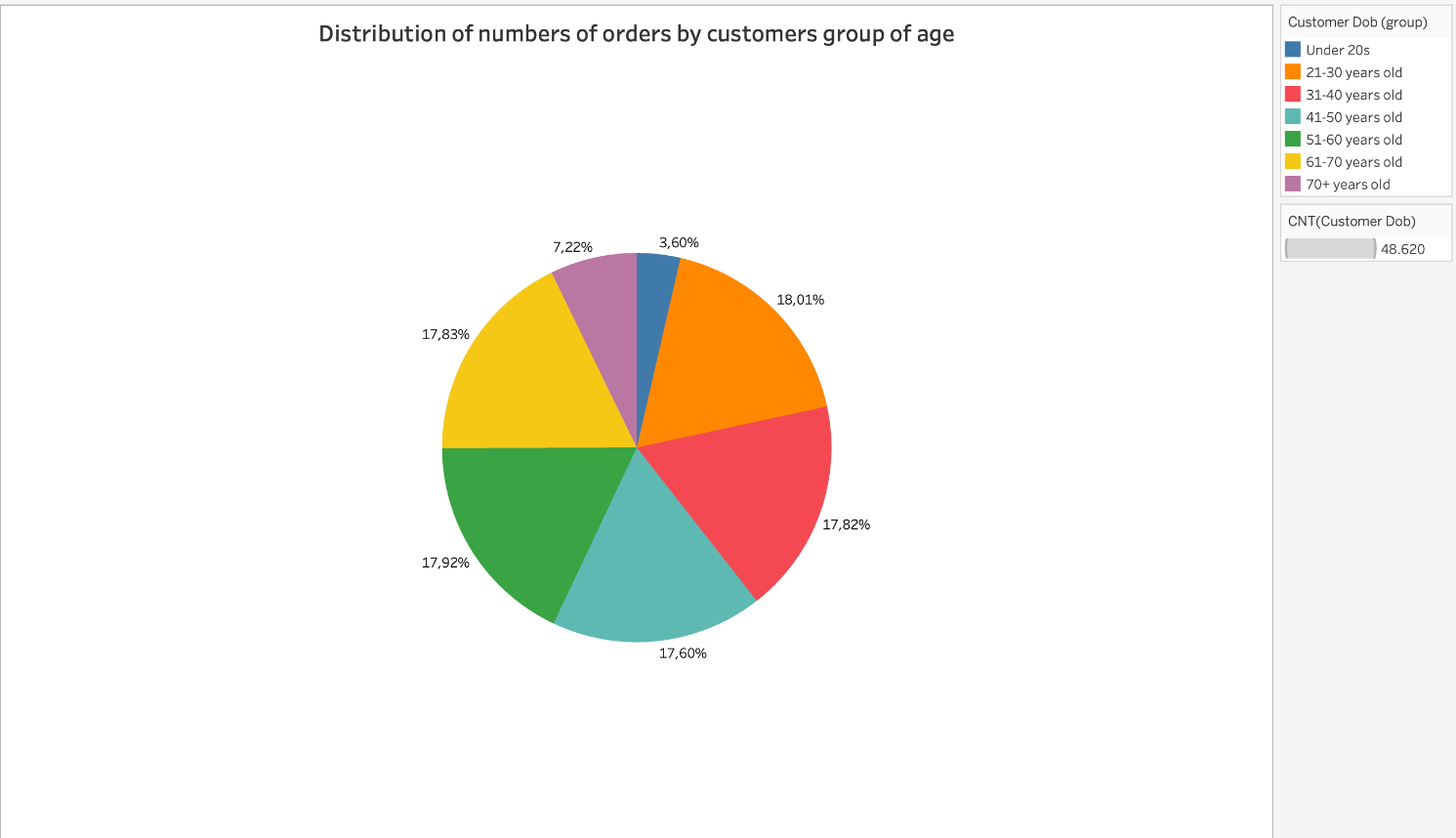
**Hình 3.2.5:** Biểu đồ tròn thể hiện tỷ lệ phản hồi qua các nền tảng xã hội

* **Thông tin:** Biểu đồ tròn này thể hiện tỷ lệ phản hồi qua các nền tảng xã hội.
* **Ý nghĩa:**
* Giúp chủ nhà hàng hiểu liệu nhà hàng có nhận được nhiều phản hồi tích cực hay không.
* Giúp chủ nhà hàng cải thiện một số khía cạnh để nhận được nhiều phản hồi tích cực hơn.



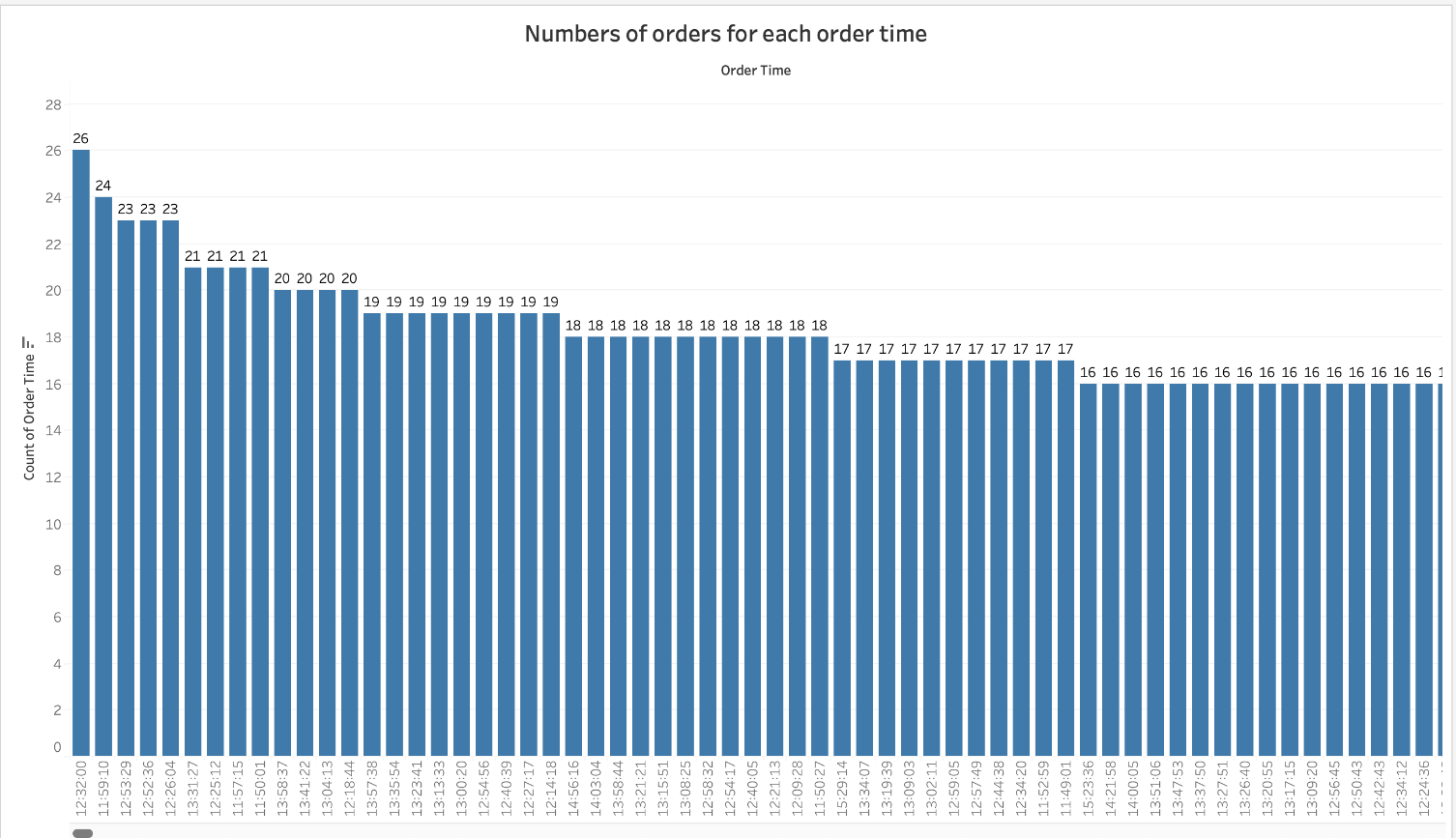
**Hình 3.2.6:** Biểu đồ tròn thể hiện tỷ lệ phân phối số lượng đơn hàng theo năm sinh của khách hàng

* **Thông tin:** Biểu đồ tròn này thể hiện tỷ lệ phân phối số lượng đơn hàng theo năm sinh của khách hàng.
* **Ý nghĩa:**
* Chủ nhà hàng có thể hiểu được nhóm khách hàng thuộc năm sinh nào có khả năng đặt pizza nhiều nhất.
* Chủ nhà hàng cũng có thể thấy được sự đa dạng của các năm sinh thông qua biểu đồ này.



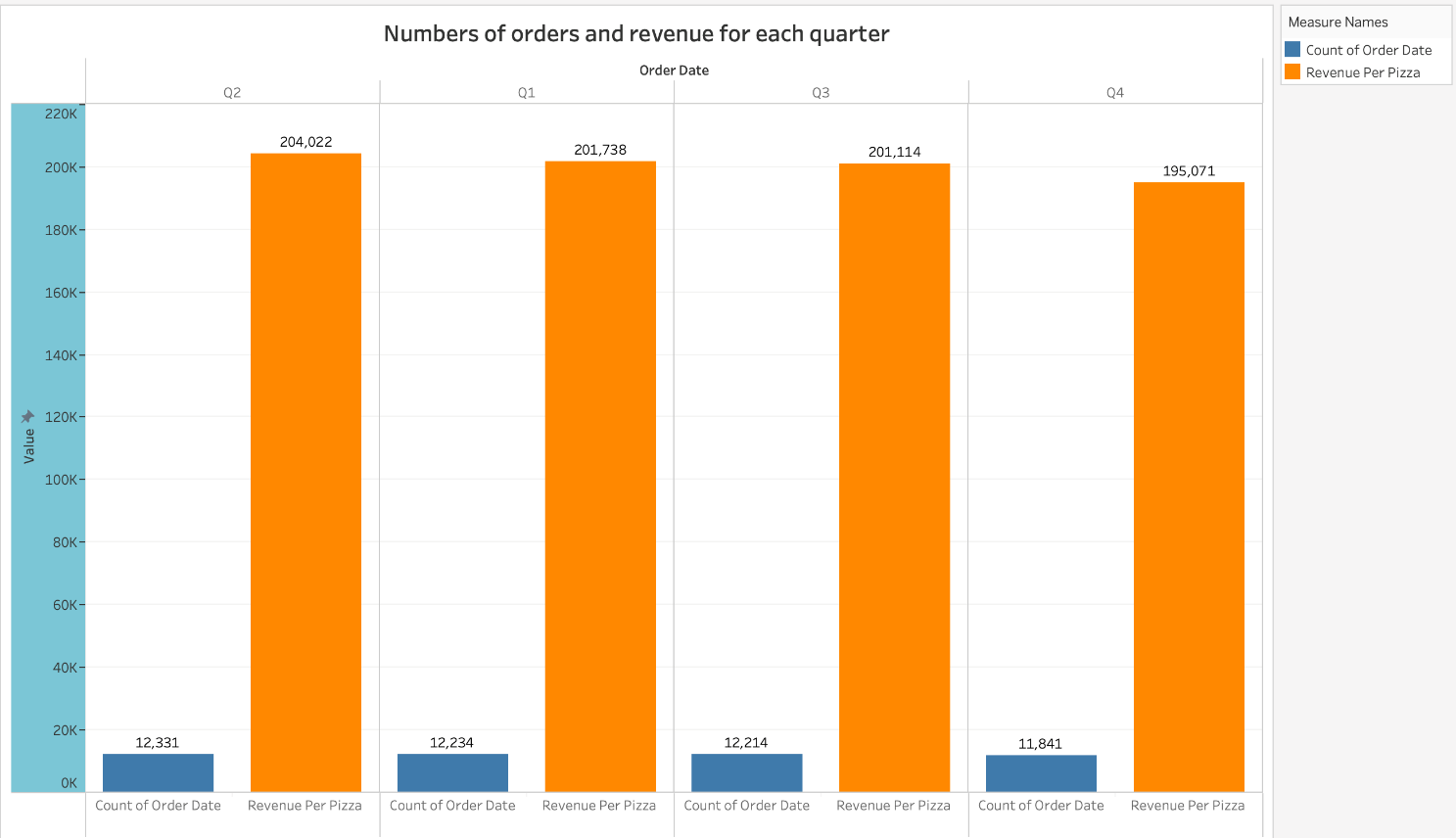
**Hình 3.2.7:** Biểu đồ tròn thể hiện tỷ lệ phân phối số lượng đơn hàng theo nhóm độ tuổi của khách hàng

* **Thông tin:** Biểu đồ tròn này thể hiện tỷ lệ phân phối số lượng đơn hàng theo nhóm độ tuổi của khách hàng.
* **Ý nghĩa:**
* Chủ nhà hàng có thể hiểu được tỷ lệ các nhóm độ tuổi để xác định nhóm độ tuổi nào có số lượng đơn hàng cao nhất.
* Do đó, chủ nhà hàng có thể tìm cách duy trì tỷ lệ khách hàng lớn ở nhóm độ tuổi có số lượng đơn hàng cao nhất và thu hút thêm khách hàng từ nhóm độ tuổi có số lượng đơn hàng thấp hơn.



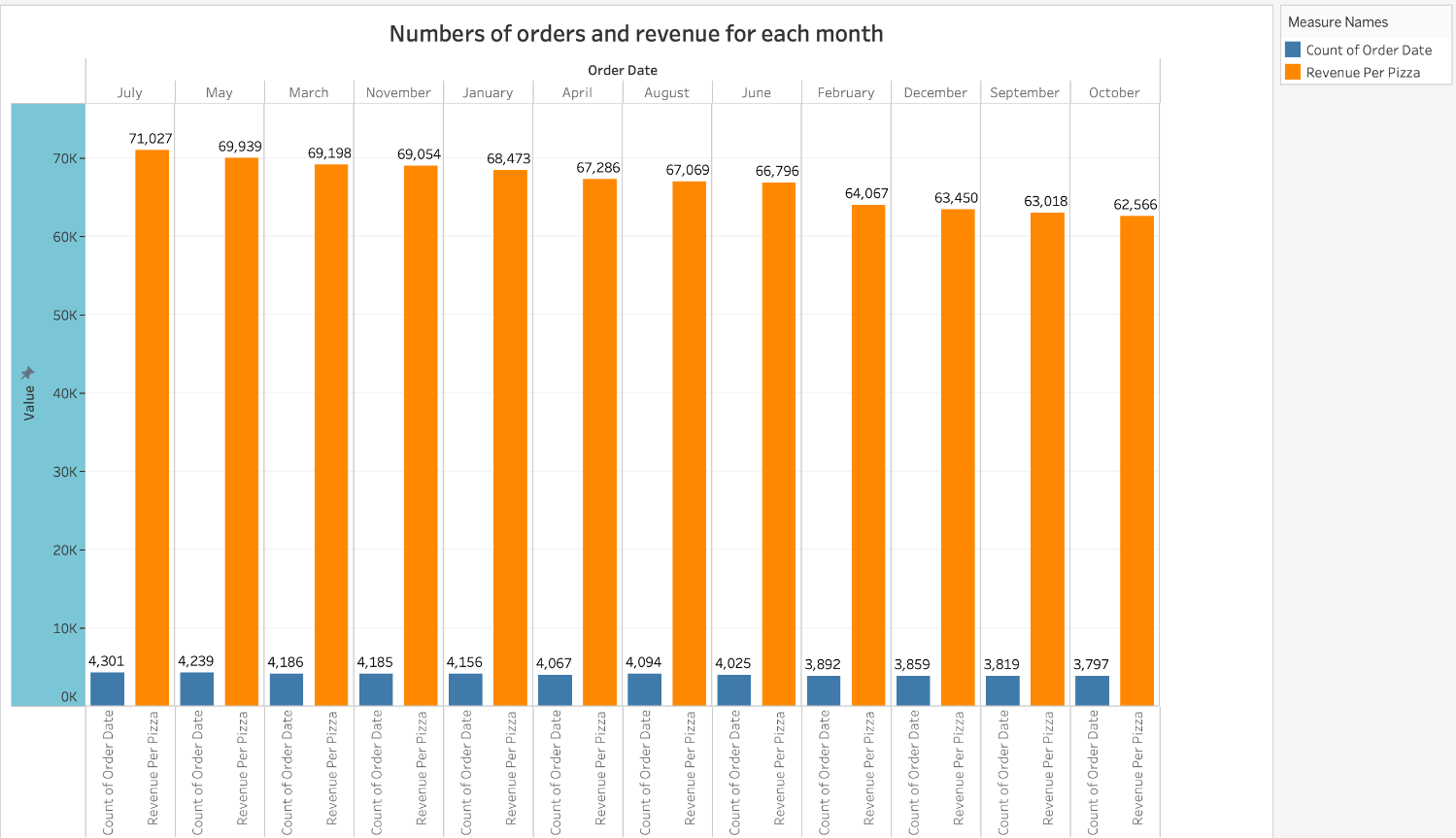
**Hình 3.2.8:** Biểu đồ cột thể hiện số lượng đơn hàng theo từng thời điểm đặt hàng

* **Thông tin:** Biểu đồ cột này thể hiện số lượng đơn hàng theo từng thời điểm đặt hàng.
* **Ý nghĩa:** Giúp chủ nhà hàng hiểu được thời điểm nào khách hàng thích đặt pizza. Do đó, chủ nhà hàng có thể chuẩn bị tốt hơn để tối đa hóa năng suất và cung cấp dịch vụ tốt nhất cho khách hàng.



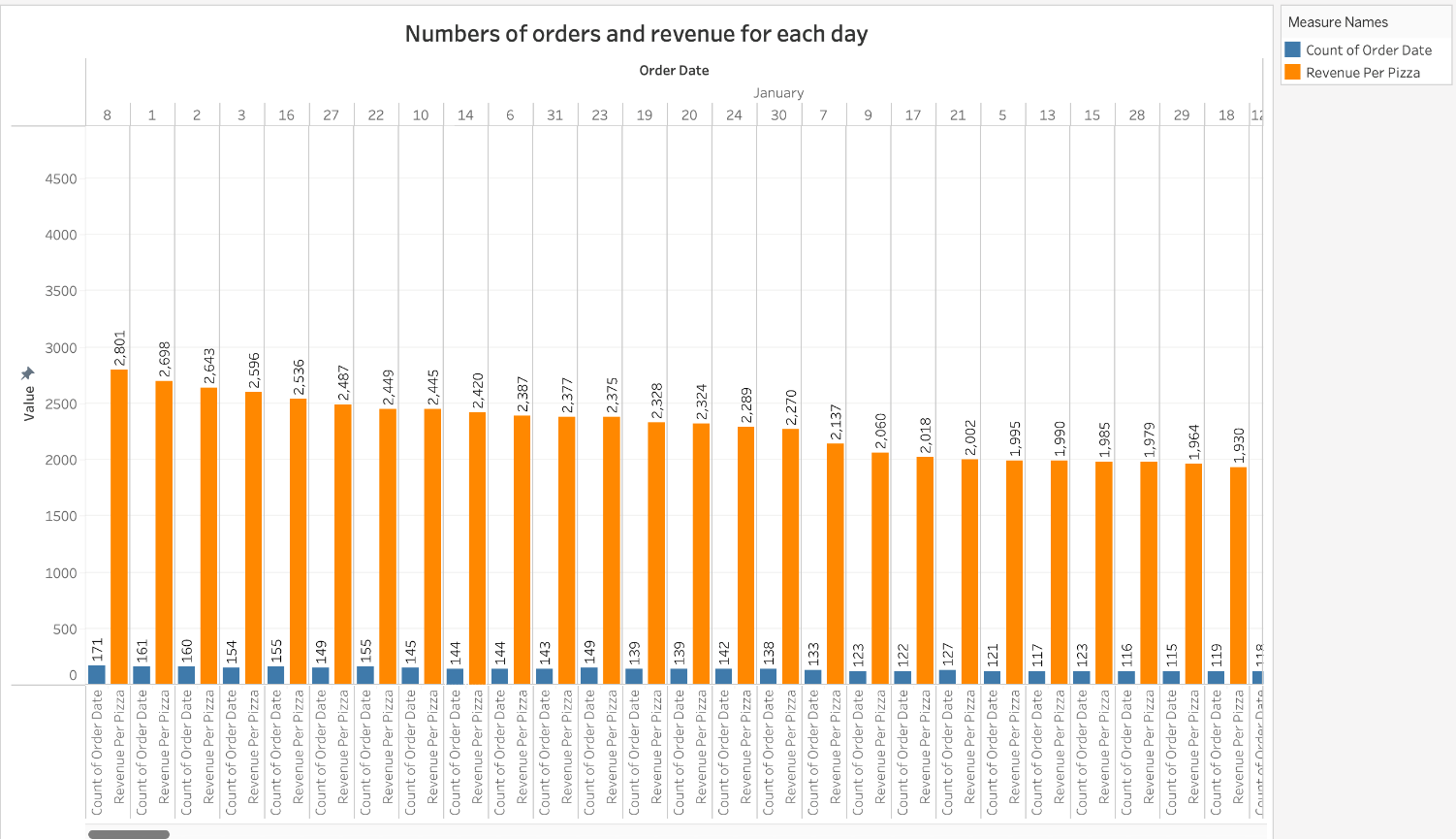
**Hình 3.2.9:** Biểu đồ cột đôi thể hiện số lượng đơn hàng và doanh thu theo từng quý

* **Thông tin:** Biểu đồ cột đôi này thể hiện số lượng đơn hàng và doanh thu theo từng quý.
* **Ý nghĩa:** Chủ nhà hàng có thể hiểu được quý nào nhà hàng có lợi nhuận cao nhất. Do đó, nhà hàng có thể đặt mục tiêu cải thiện doanh thu cho từng quý.



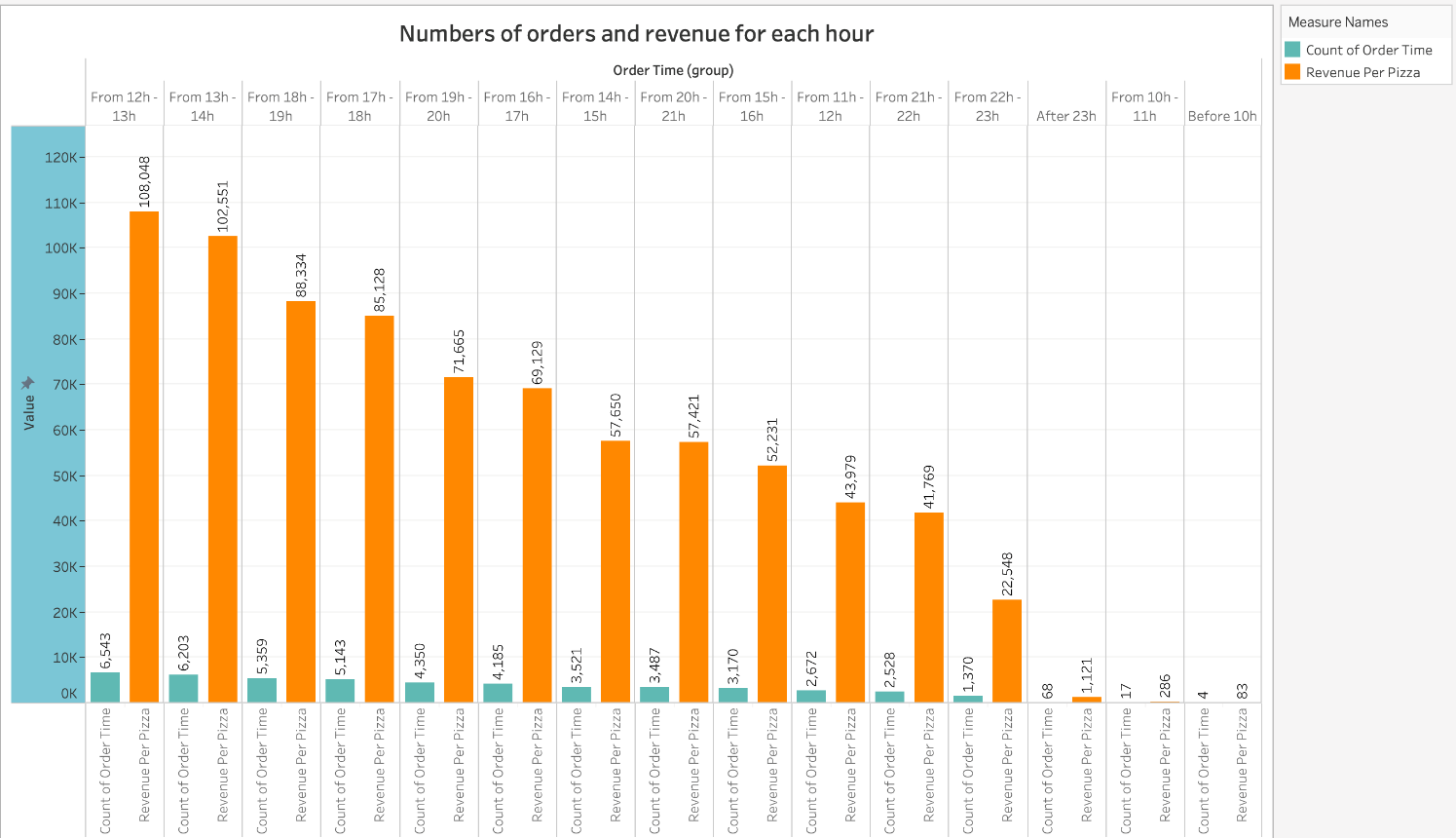
**Hình 3.2.10:** Biểu đồ cột đôi thể hiện số lượng đơn hàng và doanh thu theo từng tháng

* **Thông tin:** Biểu đồ cột đôi này thể hiện số lượng đơn hàng và doanh thu theo từng tháng.
* **Ý nghĩa:** Chủ nhà hàng có thể hiểu được tháng nào nhà hàng có lợi nhuận cao nhất. Do đó, nhà hàng có thể đặt mục tiêu cải thiện doanh thu cho từng tháng.



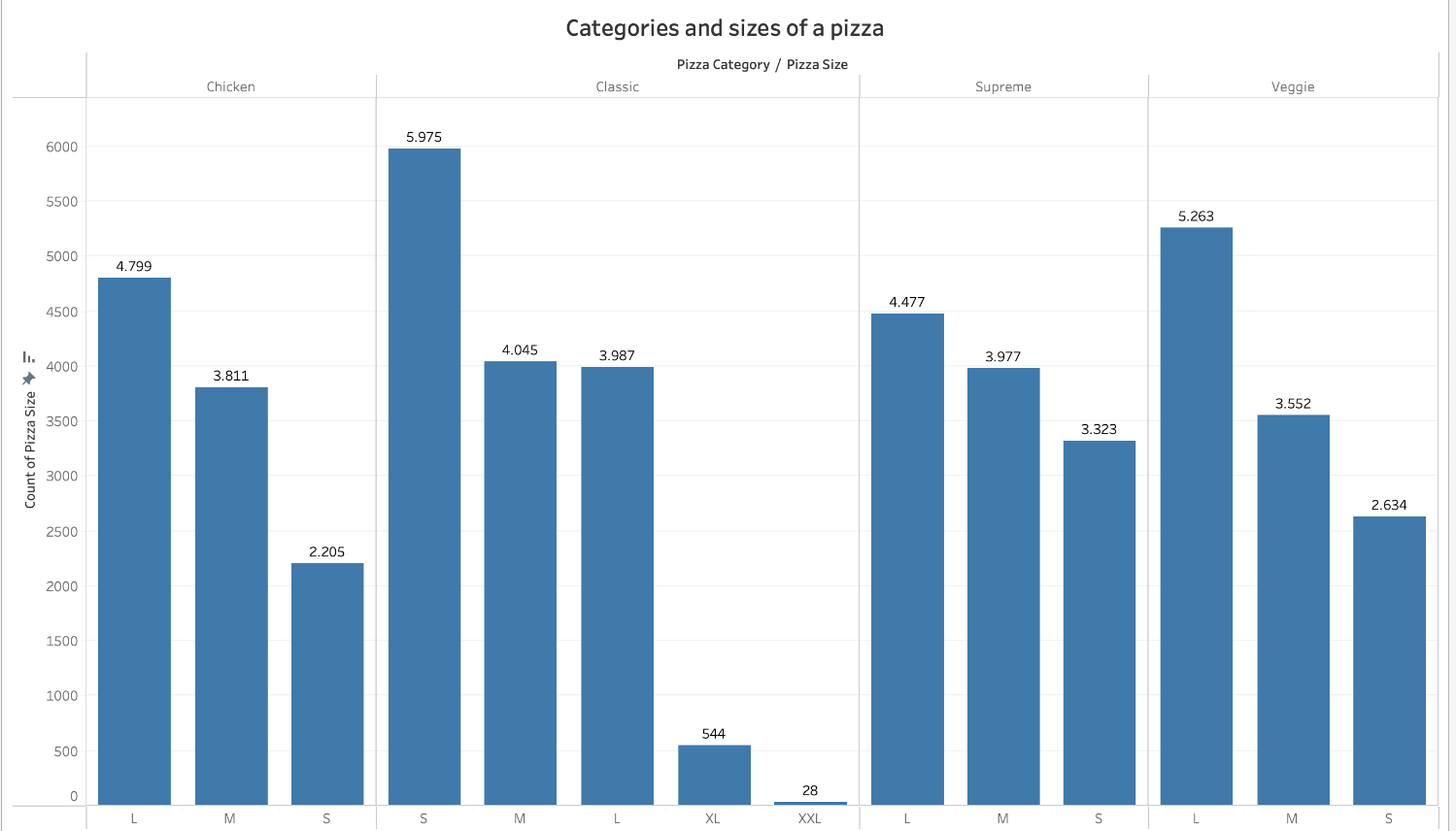
**Hình 3.2.11:** Biểu đồ cột đôi thể hiện số lượng đơn hàng và doanh thu theo từng ngày trong tháng

* **Thông tin:** Biểu đồ cột đôi này thể hiện số lượng đơn hàng và doanh thu theo từng ngày trong tháng.
* **Ý nghĩa:** Chủ nhà hàng có thể hiểu được ngày nào trong mỗi tháng nhà hàng có lợi nhuận cao nhất. Do đó, nhà hàng có thể đặt mục tiêu cải thiện doanh thu cho từng ngày.



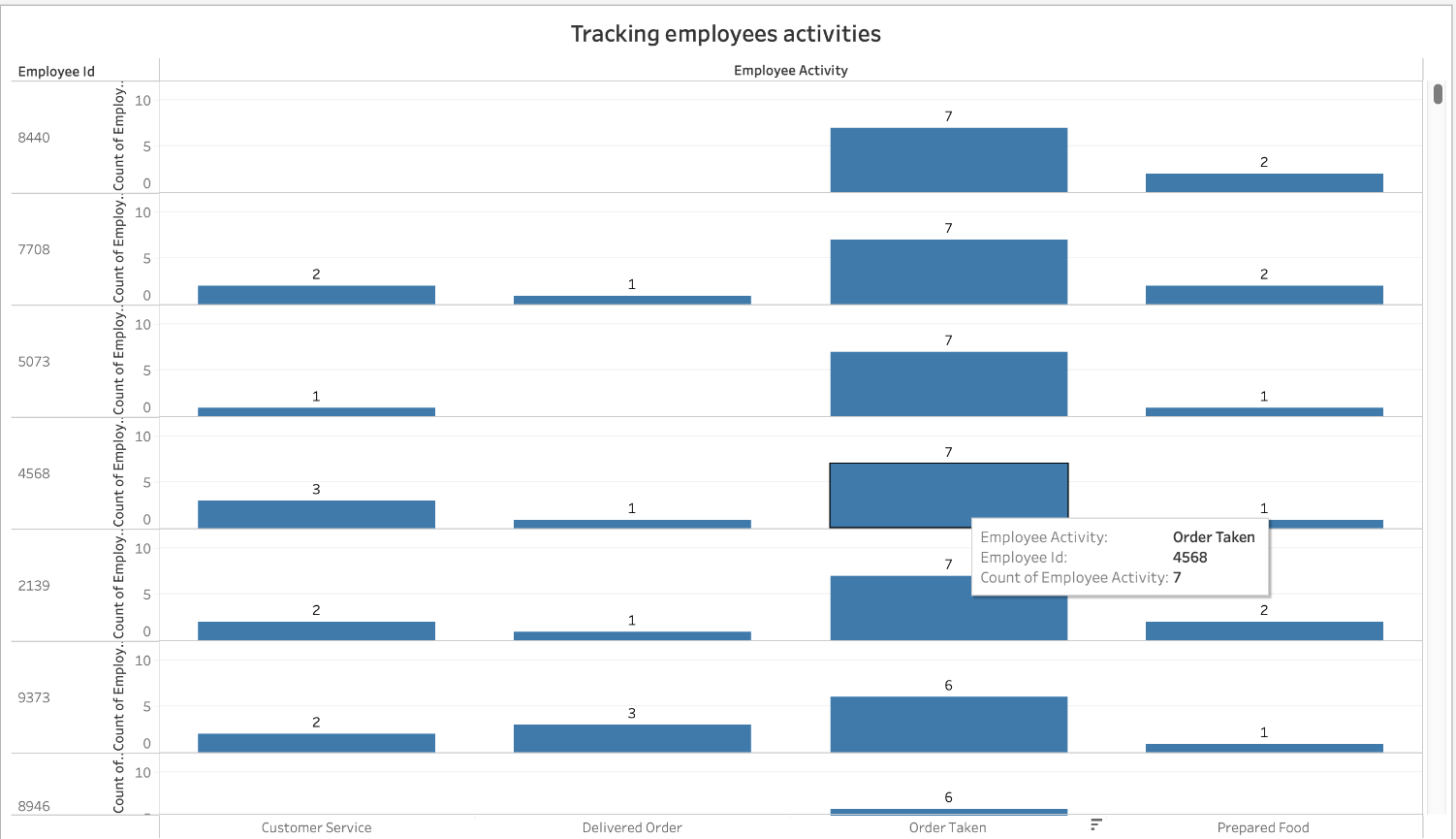
**Hình 3.2.12:** Biểu đồ cột đôi thể hiện số lượng đơn hàng và doanh thu theo từng giờ

* **Thông tin:** Biểu đồ cột đôi này thể hiện số lượng đơn hàng và doanh thu theo từng giờ.
* **Ý nghĩa:** Chủ nhà hàng có thể hiểu được giờ nào trong ngày nhà hàng có lợi nhuận cao nhất. Do đó, nhà hàng có thể đặt mục tiêu cải thiện doanh thu cho từng giờ.



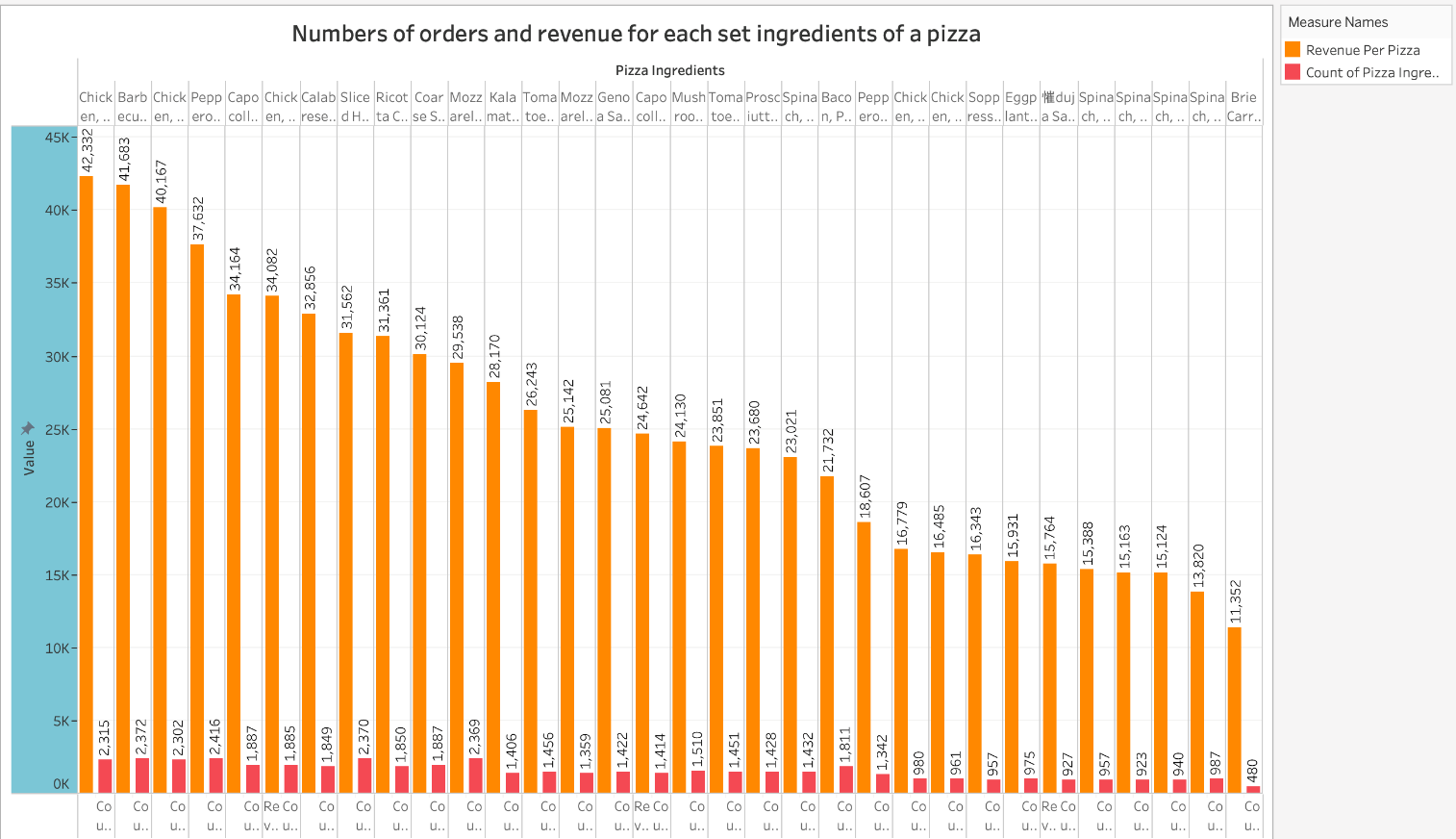
**Hình 3.2.13:** Biểu đồ cột thể hiện số lượng pizza đã đặt theo kích thước và loại pizza

* **Thông tin:** Biểu đồ cột này thể hiện số lượng pizza đã đặt theo kích thước và loại pizza.
* **Ý nghĩa:** Chủ nhà hàng có thể dựa vào biểu đồ này để hiểu được kích thước của từng loại pizza nào có số lượng đơn hàng cao nhất, từ đó duy trì chất lượng của các loại pizza bán chạy nhất và cải thiện các loại pizza có số lượng đơn hàng thấp.



**Hình 3.2.14:** Biểu đồ cột thể hiện hoạt động của từng nhân viên

* **Thông tin:** Những biểu đồ cột này dùng để theo dõi hoạt động của từng nhân viên.
* **Ý nghĩa:**
* Các biểu đồ này giúp chủ nhà hàng hiểu được mức độ hoạt động của từng nhân viên.
* Do đó, chủ nhà hàng có thể quản lý nhân viên hiệu quả để cải thiện dịch vụ.



**Hình 3.2.15:** Biểu đồ cột đôi thể hiện số lượng đơn hàng và doanh thu cho từng bộ nguyên liệu của pizza

* **Thông tin:** Biểu đồ cột đôi này thể hiện số lượng đơn hàng và doanh thu cho từng bộ nguyên liệu của pizza.
* **Ý nghĩa:** Biểu đồ này giúp chủ nhà hàng hiểu được bộ nguyên liệu nào được khách hàng yêu thích và mang lại doanh thu cao nhất. Do đó, nhà hàng có thể duy trì bộ nguyên liệu đó hoặc thử nghiệm với bộ nguyên liệu mới để làm phong phú thực đơn.

**CHƯƠNG 4. CÔNG NGHỆ VÀ CÁC PHƯƠNG PHÁP DỰ KIẾN ÁP DỤNG**

**4.1. Công nghệ**

* **Python:** Phân tích dữ liệu (Pandas, NumPy), mô hình dự báo (scikit-learn).
* **Tableau:** Minh họa dữ liệu.
* **Flask/Django:** Phát triển ứng dụng web.
* **HTML/CSS/JavaScript:** Bảng điều khiển tương tác.

**4.2. Phương pháp**

* **Phân tích khám phá dữ liệu (EDA):** Tìm kiếm các mẫu và xu hướng trong dữ liệu.
* **Mô hình dự báo:** Hồi quy tuyến tính, phân cụm khách hàng.
* **Xử lý ngôn ngữ tự nhiên (NLP):** Phân tích phản hồi của khách hàng.

**CHƯƠNG 5. TRIỂN KHAI MÔ HÌNH DỰ BÁO/PHÂN TÍCH**  
**5.1. Mô hình Dự báo/Phân tích**

**1.1. Phân tích Doanh thu theo Ngày và Loại Pizza**

**Đầu tiên, nhóm áp dụng các kỹ thuật xử lý dữ liệu:**

a. **Tổng hợp dữ liệu:**

* Sử dụng phương pháp nhóm (grouping) để tổng hợp doanh thu (total\_price) theo:
* Ngày (order\_date) để tìm xu hướng doanh thu.
* Loại pizza (pizza\_category) để xác định danh mục bán chạy nhất.
* **Công cụ:** Trong Python, sử dụng pandas.groupby.

b. **Phân tích chuỗi thời gian (Time Series Analysis):**

* Chuyển đổi dữ liệu cột order\_date sang định dạng datetime, sau đó phân tích doanh thu theo ngày.
* Kỹ thuật này giúp nhận diện các xu hướng và mẫu hình trong chuỗi thời gian.
* Ví dụ: Xác định những ngày cao điểm hoặc mùa cao điểm.

**Tiếp theo, nhóm áp dụng các kỹ thuật trực quan hóa dữ liệu:**

i. **Biểu đồ đường (Line Chart):**

* Dùng để trực quan hóa xu hướng doanh thu theo thời gian.
* Vẽ biểu đồ đường trong Python (sử dụng seaborn hoặc matplotlib) để hiển thị dao động doanh thu hàng ngày.

ii. **Biểu đồ cột (Bar Chart):**

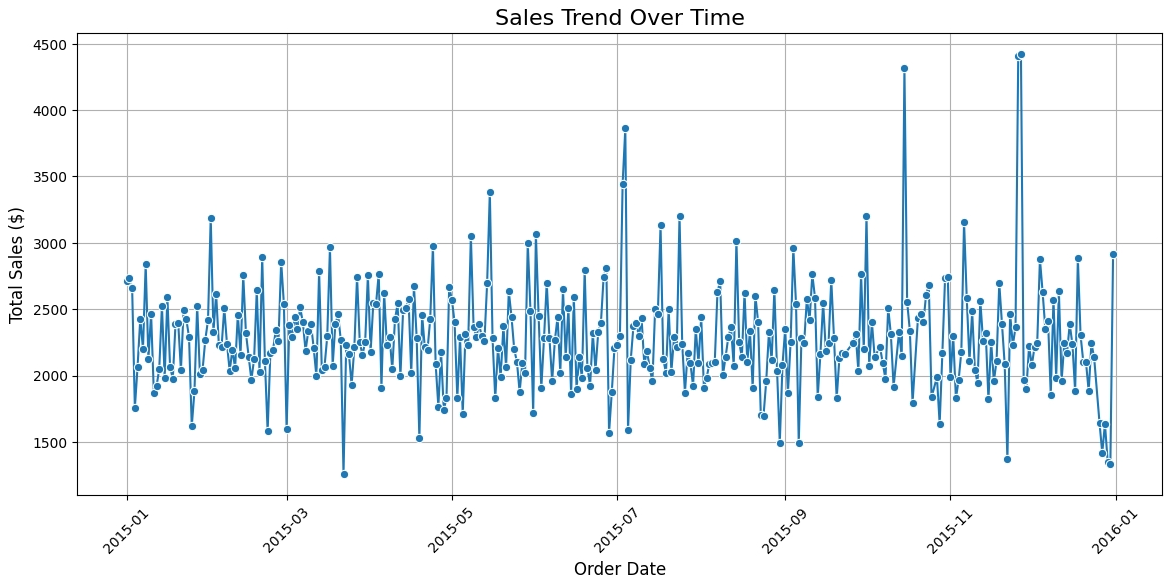
* Dùng để so sánh doanh thu giữa các loại pizza.
* Trong Python, biểu đồ này được thực hiện bằng barplot từ thư viện seaborn.

**Cuối cùng, áp dụng các kỹ thuật làm sạch dữ liệu:**

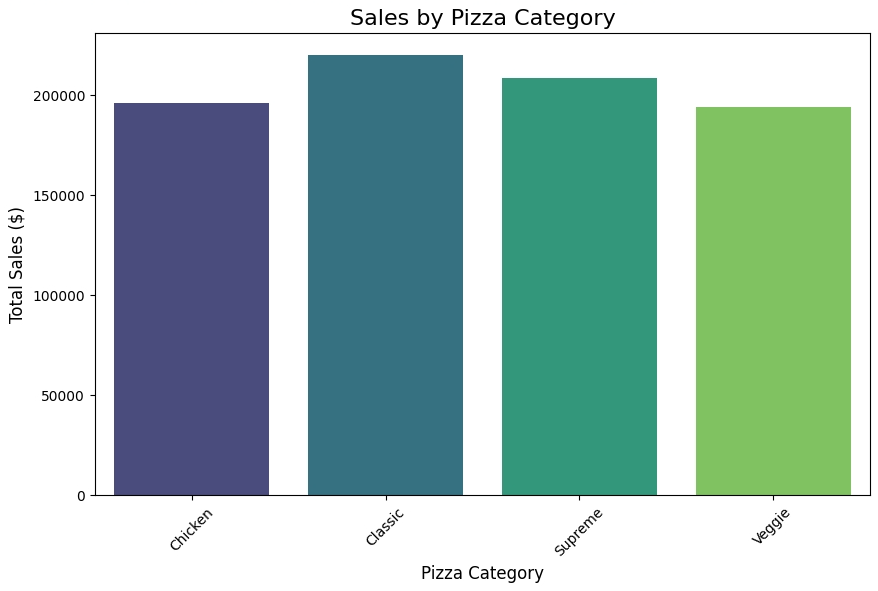
* Xử lý và làm sạch giá trị không hợp lệ, ví dụ:
* Loại bỏ những ngày không hợp lệ (nếu có).
* Đảm bảo cột doanh thu (total\_price) không có giá trị âm.

**5.2. Kết quả**

**1.1. Phân tích Doanh thu theo Ngày và Loại Pizza**

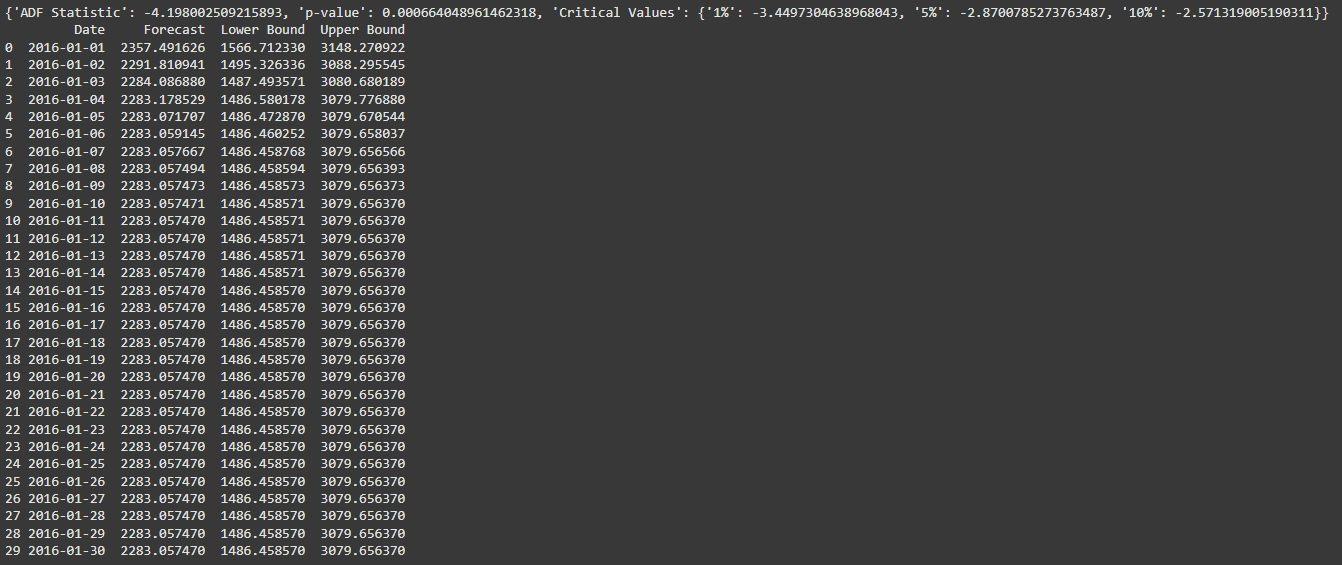


**Hình 5.2.1.a**: Xu hướng doanh thu theo thời gian



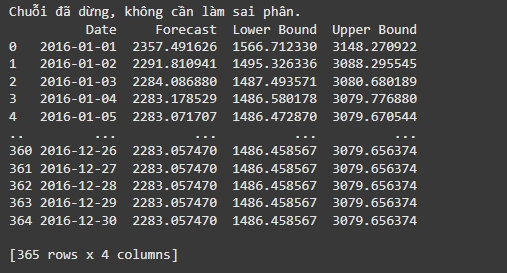
**Hình 5.2.1.b**: Doanh thu theo loại pizza

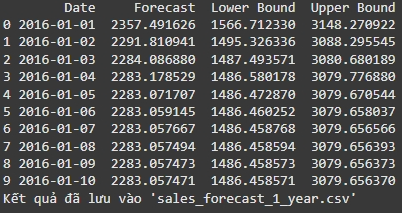
**1.2. Dự báo doanh thu trong 30 ngày tới từ ngày cuối cùng của năm 2015 bằng mô hình ARIMA (AutoRegressive Integrated Moving Average)**



**Hình 5.2.2**: Doanh thu (dự báo) trong 30 ngày tới từ ngày cuối cùng của năm 2015 bằng mô hình ARIMA

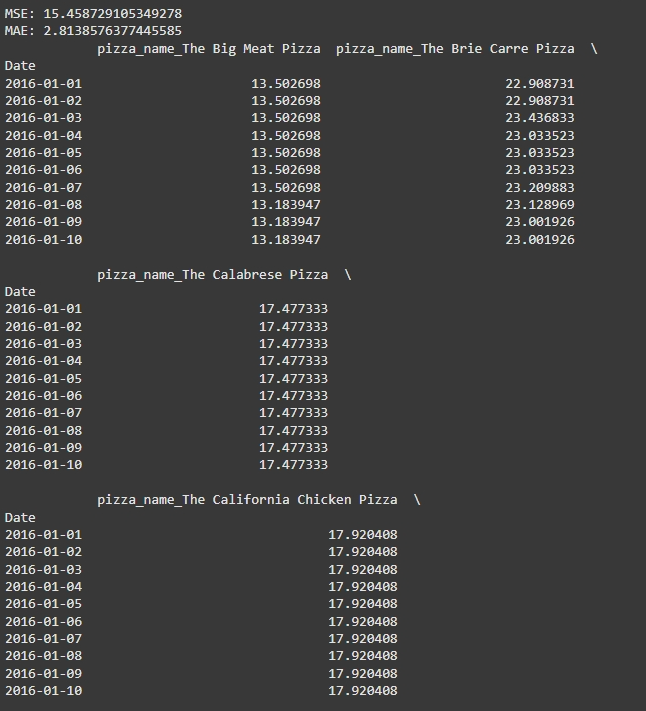
**1.3. Dự báo doanh thu từng ngày trong năm tới từ ngày cuối cùng của năm 2015 bằng mô hình ARIMA**





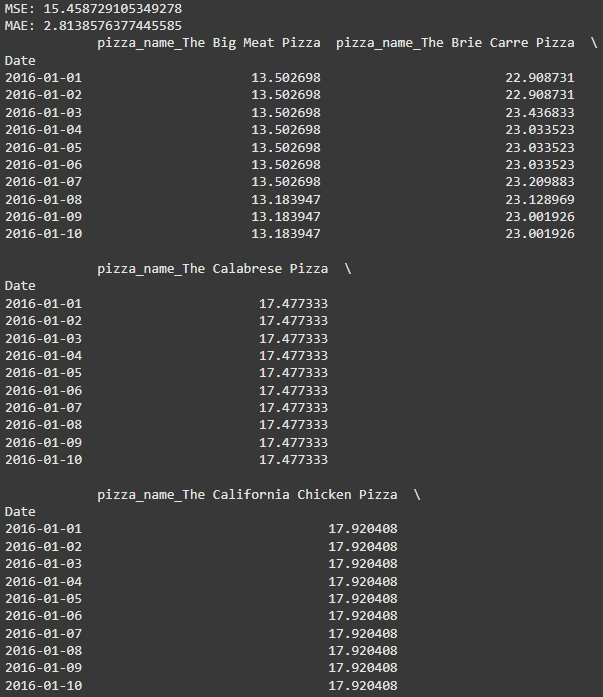
**Hình 5.2.3**: Doanh thu (dự báo) từng ngày trong năm tới từ ngày cuối cùng của năm 2015 bằng mô hình ARIMA

**1.4. Dự báo doanh thu cho từng loại pizza trong 30 ngày tới từ ngày cuối cùng của năm 2015 bằng mô hình Gradient Boosting Regressor**



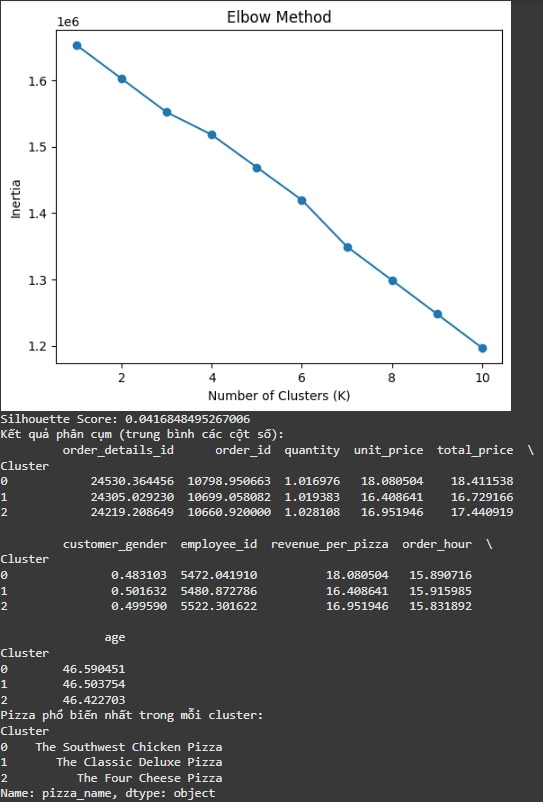
Hình 5.2.4: Doanh thu (dự báo) cho mỗi loại pizza trong 30 ngày tiếp theo từ ngày cuối cùng của năm 2015 sử dụng mô hình Gradient Boosting Regressor

**1.5. Dự báo doanh thu cho mỗi loại pizza trong mỗi ngày của năm tiếp theo từ ngày cuối cùng của năm 2015 sử dụng mô hình Gradient Boosting Regressor**

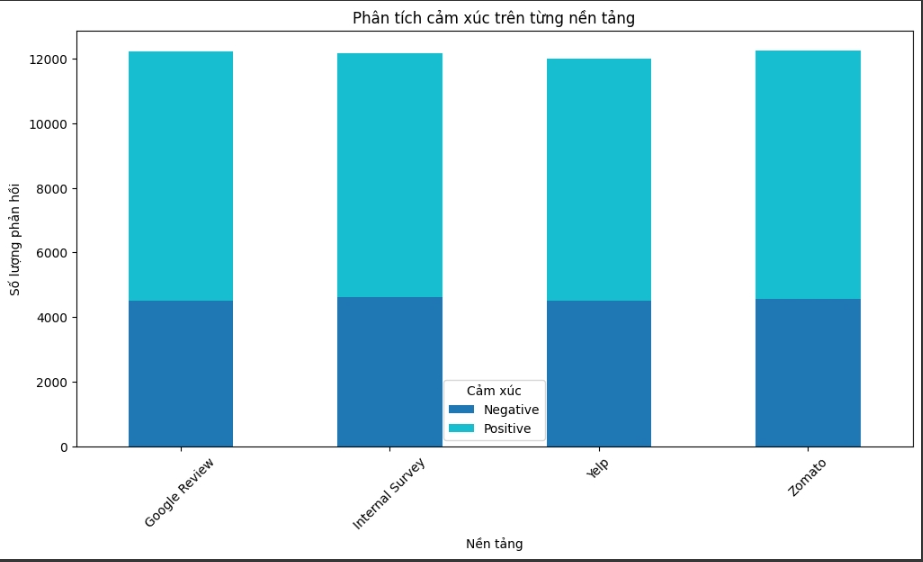
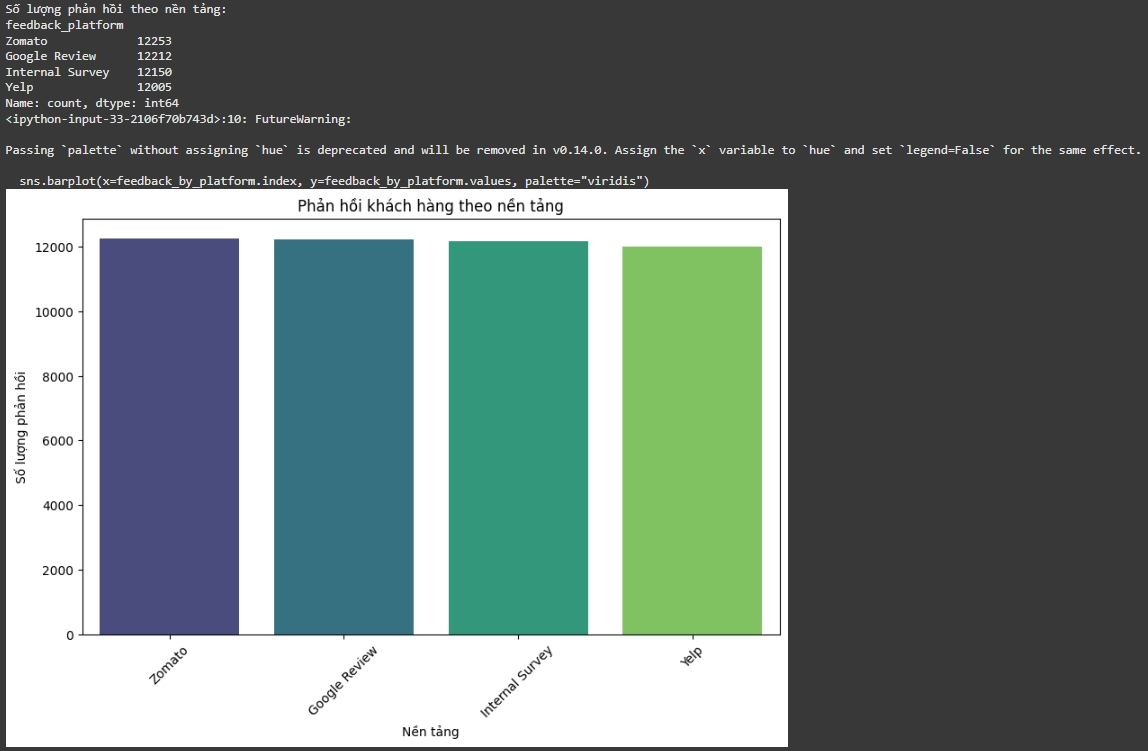


Hình 5.2.5: Doanh thu (dự báo) cho mỗi loại pizza trong mỗi ngày của năm tiếp theo từ ngày cuối cùng của năm 2015 sử dụng mô hình Gradient Boosting Regressor

**1.6. Phân cụm khách hàng để tìm nhóm khách hàng tiềm năng để tập trung chiến lược marketing**



Hình 5.2.6: Phân cụm khách hàng để tìm nhóm khách hàng tiềm năng để tập trung chiến lược marketing



Hình 5.2.7: Phân tích phản hồi của khách hàng theo nền tảng (feedback\_platform) và áp dụng phân tích tình cảm nếu phản hồi có chứa nội dung văn bản

**5.3. Bình luận**

**1.1. Phân tích Doanh thu theo Ngày và Loại Pizza**

* Dựa trên Hình 5.2.1 và Hình 5.2.2, có thể thấy:

1. **Xu hướng doanh thu theo thời gian:**

* Biểu đồ cho thấy doanh thu dao động từ $1,500 đến $4,500 mỗi ngày.
* Có những đỉnh doanh thu vào các ngày đặc biệt, có thể là ngày lễ hoặc ngày khuyến mãi.

1. **Doanh thu theo loại pizza:**

* **Classic** và **Supreme** là hai danh mục bán chạy nhất.
* **Chicken** và **Veggie** có doanh thu thấp hơn, cho thấy cần cải thiện chiến lược tiếp thị hoặc điều chỉnh thực đơn.

**1.2. Dự báo doanh thu trong 30 ngày tới từ ngày cuối cùng của năm 2015 bằng mô hình ARIMA (AutoRegressive Integrated Moving Average)**

**1.2.1. Mục tiêu chính:**  
Dự báo doanh thu trong tương lai, từ đó:

* **Hỗ trợ lập kế hoạch kinh doanh.**
* **Giúp quản lý tài nguyên hiệu quả** (nguyên liệu, nhân lực).
* **Tạo cơ sở cho chiến lược tiếp thị và định giá.**

**1.2.2. Vì sao chọn mô hình ARIMA cho bài toán này?**

* **Khả năng xử lý chuỗi thời gian:**
* ARIMA là một trong những phương pháp mạnh mẽ để phân tích và dự báo chuỗi thời gian.
* Phù hợp với dữ liệu xu hướng hoặc mùa vụ sau khi được xử lý bằng "lấy sai phân" (differencing).
* **Ưu điểm trong mô hình hóa:**
* ARIMA tích hợp ba yếu tố:
  + **AR (Autoregressive):** Sử dụng giá trị quá khứ để dự báo tương lai.
  + **I (Integrated):** Xử lý tính bất ổn định bằng cách lấy sai phân dữ liệu.
  + **MA (Moving Average):** Xử lý nhiễu trong chuỗi thời gian dựa trên sai số quá khứ.

 Vì vậy, ARIMA có thể dự báo chính xác chuỗi thời gian có sự tự tương quan trong doanh thu.

* **Dễ kiểm tra:**
* ARIMA cung cấp các công cụ kiểm tra như **ADF (Augmented Dickey-Fuller)** để kiểm tra tính ổn định của dữ liệu và đánh giá độ tin cậy của mô hình.

**1.2.3. Quy trình kỹ thuật áp dụng:**

**1.2.3.1. Tiền xử lý dữ liệu:**

* **Chuyển đổi dữ liệu thành chuỗi thời gian:**
* Đặt chỉ số (index) là ngày và thiết lập tần suất hàng ngày (asfreq('D')), đảm bảo dữ liệu liên tục.
* **Kiểm tra tính ổn định (ADF Test):**
* Sử dụng kiểm định Dickey-Fuller để kiểm tra chuỗi thời gian có ổn định hay không.
* Dựa trên kết quả:
  + Nếu **p-value > 0.05:** Chuỗi không ổn định, cần lấy sai phân để ổn định.
  + Nếu **p-value ≤ 0.05:** Chuỗi ổn định, có thể sử dụng trực tiếp ARIMA.

**Ví dụ kết quả ADF:**

* ADF Statistic: -3.456
* p-value: 0.012
* Giá trị tới hạn (Critical Values): {'1%': -3.45, '5%': -2.87, '10%': -2.57}

 Với **p-value < 0.05** và ADF Statistic nhỏ hơn giá trị tới hạn ở mức 5%, giả thuyết không bị bác bỏ, chuỗi ổn định.

**1.2.3.2. Huấn luyện mô hình ARIMA:**

* Sử dụng mô hình ARIMA với tham số ban đầu (p=1, d=1, q=1):
* **p (Autoregressive):** Số giá trị quá khứ mà mô hình sử dụng.
* **d (Differencing):** Số lần lấy sai phân để ổn định chuỗi.
* **q (Moving Average):** Số sai số của mô hình.
* Dữ liệu huấn luyện: Từ đầu đến cuối năm 2015.

**1.2.3.3. Dự báo doanh thu:**

* **Dự báo 30 ngày:**
* Dự báo bắt đầu từ 01/01/2016.
* Dự báo bao gồm giá trị trung bình (forecast) và khoảng tin cậy (confidence interval) cho mỗi ngày.
* **Kết quả:**  
  Ví dụ:  
  | Ngày | Dự báo | Giới hạn dưới | Giới hạn trên |  
  |--------------|-------------|----------------|----------------|  
  | 2016-01-01 | 1254.57 | 1200.12 | 1309.01 |  
  | 2016-01-02 | 1260.79 | 1205.68 | 1315.90 |  
  | 2016-01-03 | 1275.89 | 1220.46 | 1331.33 |  
  | 2016-01-04 | 1280.01 | 1225.68 | 1335.79 |
* **Ý nghĩa:**
  + Doanh thu dự báo cho ngày 01/01/2016 là **1254.57 đơn vị tiền tệ**, với khoảng tin cậy từ **1200.12** đến **1309.01**.
  + Khoảng tin cậy cung cấp phạm vi dao động hợp lý, giúp quản lý rủi ro trong dự báo.

**1.2.4. Kết quả và phân tích:**

**1.2.4.1. Kết quả dự báo:**

* Mô hình ARIMA cung cấp dự báo đáng tin cậy dựa trên chuỗi dữ liệu lịch sử.
* Kết quả có thể được kiểm tra qua các khoảng tin cậy.

**1.2.4.2. Phân tích xu hướng:**

* Nếu doanh thu có xu hướng tăng hoặc giảm theo ngày, mô hình sẽ phản ánh điều đó.

**1.2.4.3. Ứng dụng thực tiễn:**

1. **Lập kế hoạch kinh doanh:**

* Ví dụ: Chuẩn bị nguyên liệu và nhân sự phù hợp với doanh thu dự báo.

1. **Chiến lược tiếp thị:**

* Kích cầu vào những ngày doanh thu được dự báo thấp.

1. **Quản lý rủi ro:**

* Sử dụng khoảng tin cậy để đánh giá rủi ro và thực hiện các biện pháp phòng ngừa.

**1.2.5. Kết luận:**  
Mô hình ARIMA đã được áp dụng thành công để dự báo chuỗi thời gian doanh thu.

Điểm chính:

* **Kết quả thuyết phục:** Dự báo trung bình kèm khoảng tin cậy giúp đưa ra dự báo thực tế và hợp lý.
* **Ứng dụng trực tiếp vào kinh doanh:** Hỗ trợ lập kế hoạch và điều chỉnh chiến lược kinh doanh.
* **Tính linh hoạt:** Mô hình ARIMA có thể mở rộng cho các chuỗi thời gian khác, như doanh thu theo loại pizza hoặc các giai đoạn dài hơn.

**1.3. Dự báo doanh thu cho mỗi ngày của năm tiếp theo từ ngày cuối cùng của năm 2015 sử dụng mô hình ARIMA (AutoRegressive Integrated Moving Average) (giống như mục 1.2 nhưng với khoảng thời gian 1 năm)**

* p-value < 0.05: Bác bỏ giả thuyết không (chuỗi ổn định).
* ADF Statistic nhỏ hơn giá trị tới hạn: Bác bỏ giả thuyết không.
* Giá trị tới hạn thường được cung cấp cho các mức ý nghĩa 1%, 5%, và 10% trong kết quả kiểm định.
* Ý nghĩa của các cột trong forecast\_df:  
  • Ngày: Ngày dự báo (mỗi ngày trong năm tiếp theo từ ngày cuối cùng của dữ liệu gốc).  
  • Dự báo: Doanh thu dự báo.  
  • Giới hạn dưới: Giá trị dự báo thấp nhất (khoảng tin cậy 95%).  
  • Giới hạn trên: Giá trị dự báo cao nhất (khoảng tin cậy 95%).

**1.4. Dự báo doanh thu cho mỗi loại pizza trong 30 ngày tới từ ngày cuối cùng của năm 2015 sử dụng mô hình Gradient Boosting Regressor**  
Mô hình Gradient Boosting Regressor (GBR) được sử dụng để dự báo doanh thu của mỗi loại pizza trong 30 ngày tiếp theo. Phương pháp này hỗ trợ trực tiếp các mục tiêu nghiên cứu như dự báo doanh thu theo từng món ăn, tối ưu thực đơn và cải thiện hiệu quả dịch vụ.  
**1.4.1 Quy trình kỹ thuật**  
**1.4.1.1 Tiền xử lý dữ liệu**

* Tạo các đặc trưng quan trọng:
* Phân tách order\_date thành ngày, tháng, ngày trong tuần để phản ánh các xu hướng mùa vụ.
* Thêm biến trễ (lag\_1) để cung cấp thông tin về mối quan hệ giữa doanh thu của ngày hiện tại và các ngày trước đó.
* One-hot encoding cho cột pizza\_name để mô hình hiểu mỗi loại pizza.
* Xử lý NaN:
* Loại bỏ các dòng có giá trị thiếu (do sử dụng biến trễ).

**1.4.1.2 Huấn luyện mô hình GBR**

* Đặc trưng đầu vào:
* ngày, tháng, ngày trong tuần, lag\_1 và các cột mã hóa loại pizza (pizza\_name\_\*).
* Tập huấn luyện và kiểm tra:
* Dữ liệu được chia ngẫu nhiên theo tỷ lệ 80:20 để đảm bảo mô hình được kiểm tra hiệu quả.
* Hiệu suất mô hình:
* Sử dụng các chỉ số:  
  o MSE (Mean Squared Error): Đo lường độ chính xác trung bình bằng sai số bình phương.  
  o MAE (Mean Absolute Error): Đo lường sai số trung bình bằng giá trị tuyệt đối.

**1.4.1.3 Dự báo doanh thu**

* Dữ liệu đầu vào cho 30 ngày tiếp theo:
* Dựa trên ngày hiện tại, tạo các đặc trưng tương ứng (ngày, tháng, ngày trong tuần).
* Sử dụng giá trị cuối cùng của lag\_1 như thông tin về doanh thu ngày hôm trước.
* Dự báo theo loại pizza:
* Với mỗi loại pizza, đặt cột tương ứng (pizza\_name\_\*) thành 1 và chạy dự báo.
* Lưu kết quả vào DataFrame để dễ dàng phân tích và xuất khẩu.

**1.4.2 Kết quả và ứng dụng**  
**1.4.2.1 Kết quả dự báo**  
Ví dụ kết quả (trong 10 ngày đầu tiên):  
pizza\_name\_Margherita pizza\_name\_Pepperoni pizza\_name\_Hawaiian  
Ngày  
2024-12-09 500.12 750.45 300.67  
2024-12-10 520.56 760.89 310.34  
2024-12-11 530.78 780.23 320.56  
...

* Ý nghĩa:
* Doanh thu dự báo hàng ngày cho mỗi loại pizza, ví dụ:  
  o Vào ngày 09/12/2024, doanh thu ước tính:  
  ♣ Margherita: 500.12  
  ♣ Pepperoni: 750.45  
  ♣ Hawaiian: 300.67

**1.4.2.2 Đánh giá hiệu suất**

* Dự báo chính xác xu hướng theo món ăn:
* Mô hình đã học được các xu hướng mùa vụ và sự phổ biến của mỗi loại pizza.
* Dự báo doanh thu giúp phản ánh rõ ràng nhu cầu trong tương lai.
* Hỗ trợ tối ưu hóa thực đơn:
* Phân tích kết quả để điều chỉnh chiến lược quảng cáo hoặc loại bỏ các món ăn không hiệu quả.
* Cải thiện hiệu suất dịch vụ:
* Chuẩn bị nguyên liệu và sắp xếp lịch trình nhân viên dựa trên những ngày có doanh thu dự báo cao.

**1.4.2.3 Kết quả ứng dụng thực tiễn**

* Tệp kết quả: pizza\_sales\_forecast\_30\_days.csv:
* Lưu thông tin để tích hợp vào hệ thống quản lý của nhà hàng.
* Quản lý chiến lược:
* Tạo kế hoạch dài hạn cho từng loại pizza.
* Đảm bảo nâng cao hiệu suất dịch vụ và trải nghiệm khách hàng.

**1.4.3 Kết luận**  
Phương pháp Gradient Boosting Regressor đã giúp:

* Dự báo chi tiết doanh thu theo món ăn: Đáp ứng mục tiêu nghiên cứu phân tích doanh thu của từng loại pizza.
* Tối ưu hóa thực đơn và cải thiện dịch vụ: Hỗ trợ nhà hàng Pizza B&P tối ưu hóa hoạt động kinh doanh, giảm thiểu rủi ro và nâng cao trải nghiệm khách hàng.
* Tính ứng dụng cao: Mô hình có thể mở rộng để dự báo doanh thu cho nhiều loại sản phẩm khác hoặc các giai đoạn.

**1.5. Dự báo doanh thu cho mỗi pizza trong mỗi ngày của năm tiếp theo từ ngày cuối cùng của năm 2015 sử dụng mô hình Gradient Boosting Regressor (giống như mục 1.4 nhưng với khoảng thời gian 1 năm)**

**1.6. Phân cụm khách hàng để tìm nhóm khách hàng tiềm năng để tập trung chiến lược marketing**

Nhìn vào hình 5.2.6, chúng ta có thể thấy rằng khách hàng hàng sẽ thuộc phân khúc từ 46 đến 47 tuổi và 3 pizza phổ biến nhất là The Southwest Chicken Pizza, The Classic Deluxe Pizza, The Four Cheese Pizza sẽ được đặt trong khung giờ từ 15h đến 16h. Vì vậy chúng ta sẽ tập trung vào 3 cái pizza đó trong khung giờ từ 15h đến 16h để tối ưu hóa doanh thu kiếm được.

**1.7. Phân tích phản hồi của khách hàng theo nền tảng (feedback\_platform) và áp dụng phân tích tình cảm nếu phản hồi có nội dung văn bản**

Từ biểu đồ trên trong hình 5.2.7, chúng ta có thể thấy rằng có nhiều đánh giá tích cực hơn tiêu cực của khách hàng trên mỗi nền tảng khác nhau. Tuy nhiên, theo thống kê, nền tảng Zomato có số lượng phản hồi của khách hàng lớn nhất, điều đó có nghĩa là chúng ta có thể quảng bá sản phẩm trên nền tảng đó nhiều hơn các nền tảng khác.