VIETNAM GENERAL CONFEDERATION OF LABOUR

TON DUC THANG UNIVERSITY

FACULTY OF INFORMATION TECHNOLOGY

HO LE TRUNG TIN - 520H0585

PHAM HOANG SON - 521H0476

NGUYEN DINH VIET HOANG - 522H0120

OPTIMIZING RESTAURANT SERVICES THROUGH BUSINESS INTELLIGENCE TOOLS: A CASE STUDY OF THE PIZZA B&P RESTAURANT CHAIN

FINAL REPORT

BUSINESS INTELLIGENCE SYSTEMS

HO CHI MINH CITY, YEAR 2024

VIETNAM GENERAL CONFEDERATION OF LABOUR

TON DUC THANG UNIVERSITY

FACULTY OF INFORMATION TECHNOLOGY

HO LE TRUNG TIN - 520H0585

PHAM HOANG SON - 521H0476

NGUYEN DINH VIET HOANG - 522H0120

OPTIMIZING RESTAURANT SERVICES THROUGH BUSINESS INTELLIGENCE TOOLS: A CASE STUDY OF THE PIZZA B&P RESTAURANT CHAIN

FINAL REPORT
BUSINESS INTELLIGENCE SYSTEMS

Advised by

MsC. DUONG HUU PHUC

HO CHI MINH CITY, YEAR 2024

ACKNOWLEDGEMENT

We sincerely thank MsC. Duong Huu Phuc for teaching us the Business Intelligence Systems course with great enthusiasm. We want to express our deep appreciation for the dedication and professional knowledge that you shared with us. Through your classes, we gained a better understanding of the fundamental aspects of the Business Intelligence Systems, thanks to your detailed explanations and practical applications. You helped us grasp the knowledge and apply it effectively. Finally, we extend our heartfelt gratitude to MsC. Duong Huu Phuc for your commitment and invaluable support throughout our learning journey in this course. The skills and knowledge we acquired will continue to impact our future development. We sincerely thank you and wish your health, success, and happiness.

Ho Chi Minh City, November 14, 2024

Authors:

Ho Le Trung Tin

Pham Hoang Son

Nguyen Dinh Viet Hoang

DECLARATION OF AUTHORSHIP

We hereby declare that this thesis was carried out by ourselves under the guidance and supervision of MsC. Duong Huu Phuc; and that the work and the results contained in it are original and have not been submitted anywhere for any previous purposes. The data and figures presented in this thesis are for analysis, comments, and evaluations from various resources by our own work and have been duly acknowledged in the reference part.

In addition, other comments, reviews and data used by other authors, and organizations have been acknowledged, and explicitly cited.

We will take full responsibility for any fraud detected in my thesis. Ton Duc Thang University is unrelated to any copyright infringement caused on my work (if any).

Ho Chi Minh City, November 14, 2024

Authors:

Ho Le Trung Tin

Pham Hoang Son

Nguyen Dinh Viet Hoang

OPTIMIZING RESTAURANT SERVICES THROUGH BUSINESS INTELLIGENCE TOOLS: A CASE STUDY OF THE PIZZA B&P RESTAURANT CHAIN

ABSTRACT

The restaurant industry faces increasing challenges in improving operational efficiency and customer satisfaction. This study explores how Business Intelligence (BI) tools can optimize restaurant services, focusing on the Pizza B&P restaurant chain as a case study. The primary objectives are to enhance menu optimization, increase sales, and improve customer experience by analyzing sales data, customer feedback, and employee performance over six months.

Key datasets include sales records, customer demographics, feedback ratings, and employee activity logs. Using tools like Python, Tableau, and Flask/Django, the project applies exploratory data analysis (EDA), predictive modeling (e.g., linear regression for sales forecasting), and customer segmentation through clustering. Additionally, natural language processing (NLP) techniques are employed to analyze customer feedback.

The study delivers actionable insights through detailed analytics reports and an interactive BI dashboard. The dashboard provides real-time visualization of sales trends, customer segmentation, and staff performance. Forecasting models identify best-selling dishes and customer groups, enabling strategic recommendations such as targeted promotions and service time optimization. A web application integrates these features, empowering restaurant managers to make data-driven decisions and enhance operational efficiency.

By implementing these BI-driven strategies, the study demonstrates the potential of leveraging data analytics to improve restaurant performance and customer satisfaction in the competitive foodservice industry.

TABLE OF CONTENTS

LIST OF FIGURES..ix

CHAPTER 1. INTRODUCTION TO THE TOPIC..1

1.1. Topic 1

1.2. Objective..1

1.3. Reason for choosing the topic...2

1.4. Scope 2

1.5. Output..2

CHAPTER 2. INTRODUCTION TO DATASET..3

2.1. Data specification...3

2.2. Data features relationship..3

2.3. Why choose this dataset over another? 4

CHAPTER 3. DATA VISUALIZATION AND DATA PREPROCESSING.................7

3.1. Preprocessing 7

3.2. Visualization...7

CHAPTER 4. TECHNOLOGY AND APPROACHES EXPECTED TO USE 23

4.1. Technology 23

4.2. Methods 23

CHAPTER 5. IMPLEMENTING FORECASTING/ANALYTICAL MODELS 24

5.1. Forecasting/Analytical model...24

5.2. Result 25

5.3. Comment 33

CHAPTER 6. BUILD AN APP WITH A DASHBOARD 42

REFERENCES 43

LIST OF FIGURES

Figure 3.2.1: Best-selling pizzas in the scale of 10 items chart 8
Figure 3.2.2: Best-selling pizzas in the scale of 100 items chart 9
Figure 3.2.3: Pizza ordering frequency of customers 10
Figure 3.2.4: Bar chart illustrates the number of each feedbacks through social platforms 11
Figure 3.2.5: Pie chart showing the ratio of feedbacks through social platforms..........12
Figure 3.2.6: Pie chart shows the ratio distribution of number of orders by customers brith year 13
Figure 3.2.7: Pie chart shows the ratio distribution of number of orders by customers group of age..14
Figure 3.2.8: Bar chart shows the numbers of orders for each order time.....................15
Figure 3.2.9: Double-bar chart displays numbers of orders and revenue for each quarter 16
Figure 3.2.10: Double-bar chart displays numbers of orders and revenue for each month 17
Figure 3.2.11: Double-bar chart displays numbers of orders and revenue for each day from each month 18
Figure 3.2.12: Double-bar chart displays numbers of orders and revenue for each hour 19
Figure 3.2.13: Bar chart displays the number of ordered pizzas based on sizes and categories...20
Figure 3.2.14: Bar charts for tracking each employee activities....................................21
Figure 3.2.15: Double-bar chart shows the numbers of orders and revenue for each set ingredients of a pizza 22
Figure 5.2.1.a: Sales Trend Over Time...25
Figure 5.2.1.b: Sales by Pizza Category 26

Figure 5.2.2: Revenue (forecasted) for the next 30 days from the end of the last day of 2015 using ARIMA model....27
Figure 5.2.3: Revenue (forecasted) for each day of the next year from the end of the last day of 2015 using ARIMA model....28
Figure 5.2.4: Revenue (forecasted) for each pizza for the next 30 days from the end of the last day of 2015 using Gradient Boosting Regressor model....29
Figure 5.2.5: Revenue (forecasted) for each pizza for each day of the next year from the end of the last day of 2015 using Gradient Boosting Regressor model....30
Figure 5.2.6: Cluster customers to find potential customer groups to focus marketing strategies....31
Figure 5.2.7: Analyze customer feedback by platform (feedback_platform) and apply sentiment analysis if the feedback contains text content 33

CHAPTER 1. INTRODUCTION TO THE TOPIC

1.1. Topic

Optimizing restaurant services through Business Intelligence: A case study of Pizza B&P restaurant chain.

1.2. Objective

Optimize the business operations of Pizza B&P restaurant chain through BI tools by:

- Predicting sales by dish.
    - Meaning:
        - Understanding the consumption trend of each pizza type in the next 30 days/1 year.
        - Supporting the management of raw material supply, especially for fast-selling items.
    - Practical application:
        - Adjusting the quantity of imported ingredients to avoid waste and shortage.
        - Planning promotions for slow-selling pizzas to boost sales.
- Optimizing the menu based on sales data analysis and customer feedback.
    - Meaning:
        - Analyzing forecast data to identify fast-selling and slow-selling items.
        - Using the information to improve the menu, focusing on popular items.
    - Practical application:
        - Eliminate or adjust the prices of slow-selling items.
        - Increase the variety of best-selling dishes or create new dishes based on trends.

- Improving service efficiency, reducing service time.
    - Meaning:
        - Forecasting daily orders helps optimize staff schedules.
        - Reduce service time by preparing popular dishes in advance.
    - Practical application:
        - Increase staffing on days when demand is forecast to be high.
        - Arrange ingredients properly to reduce preparation time.

1.3. Reason for choosing the topic

Many restaurant chains have difficulty in:

- Predicting customer demand and optimizing resources.
- Understanding customer feedback to improve service.
- Increasing sales of low-performing dishes.

⟹ Our team decided to use real-world data from the B&P restaurant chain to directly support the above objectives, while providing rich information to deploy BI solutions.

1.4. Scope

Analyzing data from Pizza B&P restaurant chain for 12 months.

Including sales data for each dish, customer feedback.

1.5. Output

1. Detailed reports with analytics and visualizations.
2. BI dashboard displaying sales and customer feedback.
3. Sales prediction and customer segmentation models.
4. Suggesting strategies to improve restaurant performance.

CHAPTER 2. INTRODUCTION TO DATASET

2.1. Data specification

The Expanded_Pizza_Sales.xlsx file includes 48620 rows and 20 columns, specifically:

1. Order information:
   - order_details_id, order_id, pizza_id, quantity
2. Order time:
   - order_date, order_time
3. Pizza price and type information:
   - unit_price, total_price, pizza_size, pizza_category
4. Ingredients and dish name:
   - pizza_ingredients, pizza_name
5. Customer information:
   - customer_name, customer_gender, customer_dob
6. Customer feedback:
   - customer_feedback, feedback_platform, feedback_date
7. Employee activity:
   - employee_id, employee_activity

2.2. Data features relationship

Relationship between features in dataset:

1. Order data:
   - order_date: Range from 2015-01-01 to 2015-12-31.
   - order_time: There are 16,382 different values, showing very detailed data about time.
   - quantity (average): ~1.02, most customers order 1 pizza/time.
2. Price information:
   - unit_price: From $9.75 to $35.95, average $16.49.
   - total_price: The largest value is $83, probably due to large orders.

3. Pizza details:
    - There are 32 types of pizza, divided into 4 main categories and 5 different sizes.
    - Best-selling dish: The Classic Deluxe Pizza.
4. Customer feedback:
    - There are a total of 64 customers with diverse feedback, for example: "Loved the ambiance."
    - customer_gender: Has 3 value groups: Male, Female, and the “other” or unspecified group.
5. Employee activities:
    - Includes tasks such as: “Order Taken”, “Delivered Order”, “Prepared Food” and “Customer Service”.

2.3. Why choose this dataset over another?

1. Suitable for the research objective

   + This dataset provides comprehensive data, directly related to the research objective to:

    - Predict sales: Through data fields such as order_date, order_time, unit_price, total_price, and quantity.
    - Analyze customer behavior: With the customer_feedback, customer_gender, customer_dob, and pizza_preferences fields, we can classify customers, analyze ordering habits and feedback.
    - Optimize menu: Based on sales information (quantity) and dish ingredients (pizza_ingredients), we can analyze low-efficiency dishes and optimize the portfolio.
2. High level of detail and practicality

   + This dataset records 48,620 real transactions within 1 year, helping to ensure:

    - Comprehensiveness: Covers all times of the year (peak and low seasons).

- Trend analysis: Allows to identify revenue fluctuations over time (day/hour) and by branch.

+ The data reflects the actual business operations of the Pizza B&P restaurant chain, instead of a simulated data set.

3. Diversity of information

+ This dataset contains not only sales information but also includes:

- Customer information: Supports analysis of target customer groups based on age, gender, and behavior.
- Customer feedback: Provides qualitative data (feedback, sentiment) to assess satisfaction and improve service.
- Employee activities: Can be used to evaluate work performance and optimize service time.

4. Potential for exploitation with BI techniques

+ This dataset is suitable for applying advanced data processing and analysis techniques:

- Data visualization:
  - Tableau can display sales trends, customer groups, and service performance.
- Advanced Analytics:
  - Sales Forecasting: Using Regression Models.
  - Customer Clustering: Classifying Customers by Behavior and Needs.
  - Sentiment Analysis: Using Customer Feedback to Determine Satisfaction and Improve Service.

5. Practical Applications

+ The dataset is not only suitable for theoretical research but also helps deploy practical applications:

- Create Dashboards: Help managers monitor restaurant performance in real time.
- Decision Support:
    - Adjust the menu to increase sales.
    - Focus on improving service at branches with negative feedback.

6. Comparison with Other Datasets

+ Compared to other datasets (say, non-detailed sales data or simulated data), this dataset stands out because:

- More detail: Includes time, feedback, and staff – information that other datasets often do not have.
- High integration: Easy to use with Tableau, Python, or other tools for analysis and visualization.
- Reliability: Real-world data from a restaurant chain increases reliability and applicability.

CHAPTER 3. DATA VISUALIZATION AND DATA PREPROCESSING

3.1. Preprocessing

Data Preprocessing Steps:

1. Check and Handle Missing Data

   + Identify columns with missing values and decide:

     - Remove columns (if too many missing values).
     - Fill in replacement values (mean, median, or mode).

2. Handle Invalid Data

   + Check for unusual values, such as:

     - Negative values in revenue or quantity columns.
     - Invalid dates or dates outside the desired time range.

3. Standardize Format

   + Convert date columns (order_date and feedback_date) to standard datetime format.

   + Standardize category data, such as pizza size (pizza_size) or pizza category (pizza_category).

4. Create additional value columns

   + Average revenue per pizza: revenue_per_pizza = total_price / quantity.

   + Time of Day: Create an order_hour column to analyze the time of order placement.

3.2. Visualization

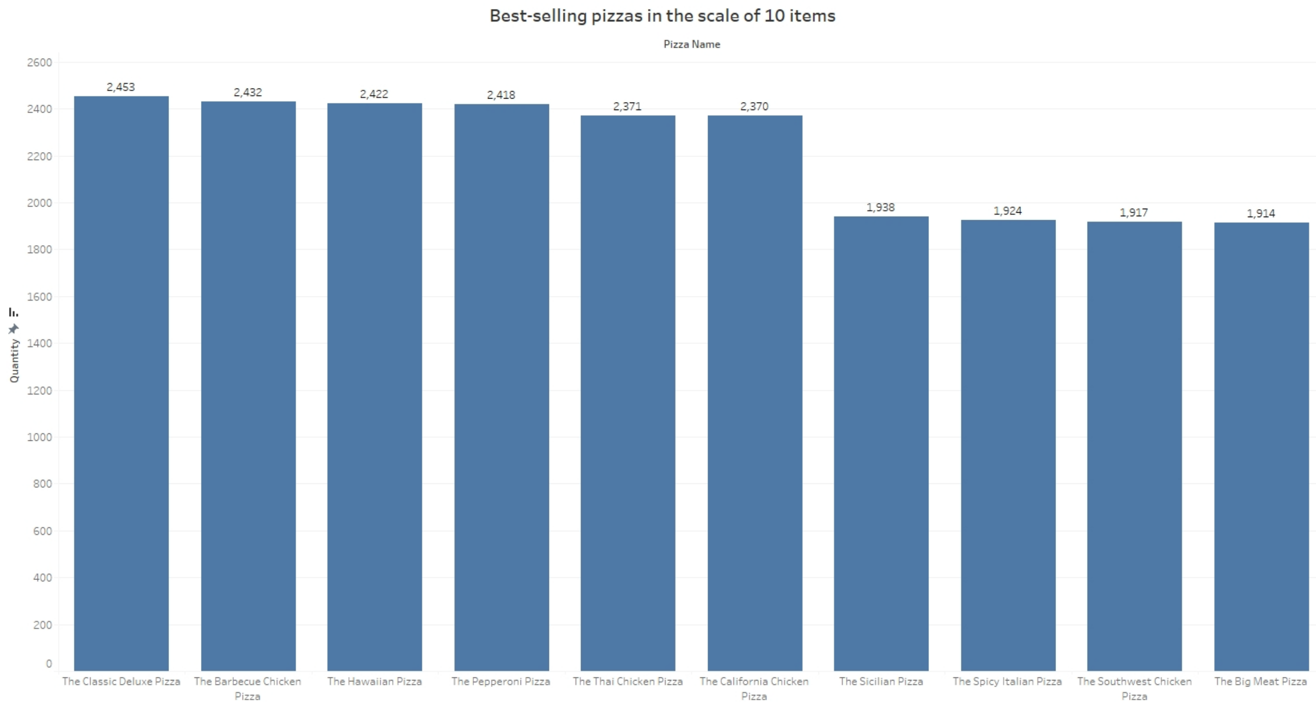


Figure 3.2.1: Best-selling pizzas in the scale of 10 items chart

+ Insights: This chart shows top 10 best-selling pizzas by quantity

+ Meaning:

- Help the restaurant understand which pizza is interested by everyone. Therefore, the restaurant know that they should maintain the quality of the pizza.

Moreover, the restaurant can have the decision to promote the pizza by giving the pizza discount on special days.

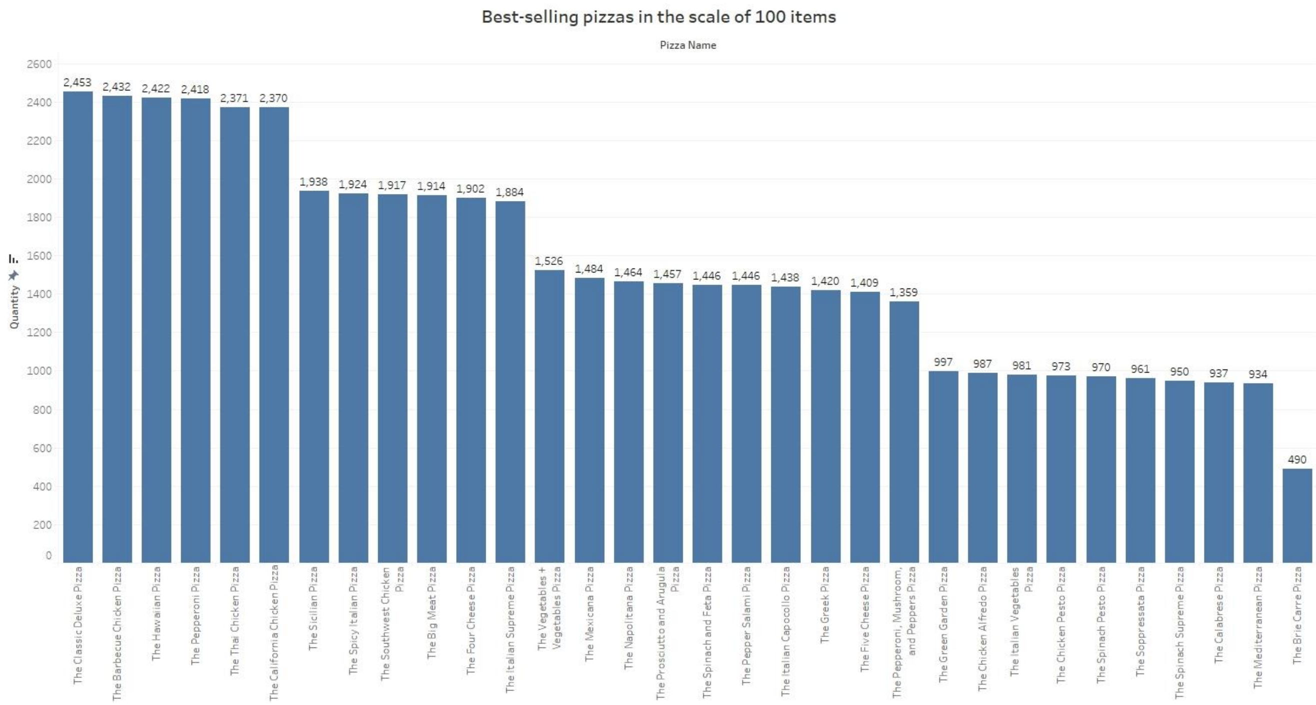


Figure 3.2.2: Best-selling pizzas in the scale of 100 items chart

+ Insights: This chart shows top 100 best-selling pizzas by quantity

+ Meaning:

- Help the restaurant understand which pizza is interested by everyone. Therefore, the restaurant know that they should maintain the quality of the pizza has the high sales number and improve the quality of the pizza has the low sales number.
- With the pizza has low sales number, the restaurant owner understand that they need to action, whether start a promotion program to improve the sales number and the revenue of that pizza or alter the pizza price to attract the customers to order it.

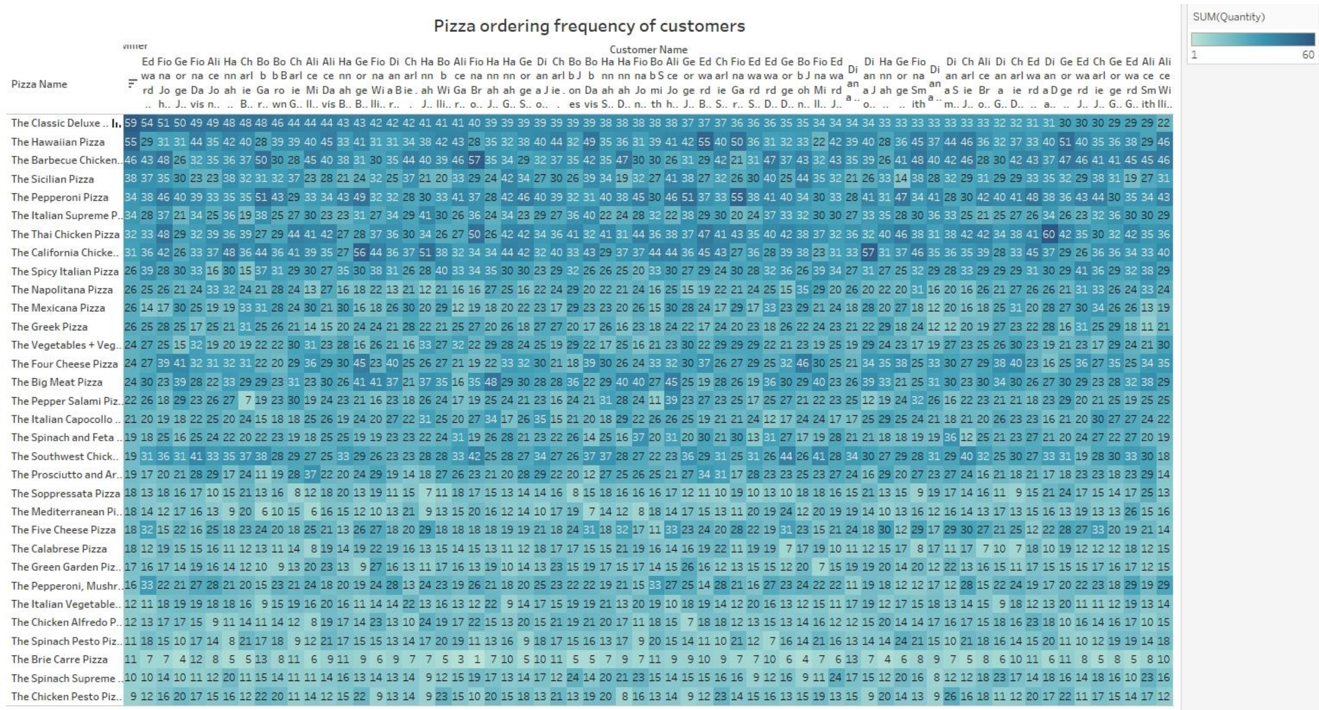


Figure 3.2.3: Pizza ordering frequency of customers

+ Insights: This figure represents the pizza ordering frequency of customers for all pizzas in the store.

+ Meaning:

- Through this figure, the restaurant owner can understand the frequency ordering each pizza of all the customers to understand customer better.
- Therefore, the restaurant can prepare some vouchers or special treatment for each customer with the pizza they ordered the most.

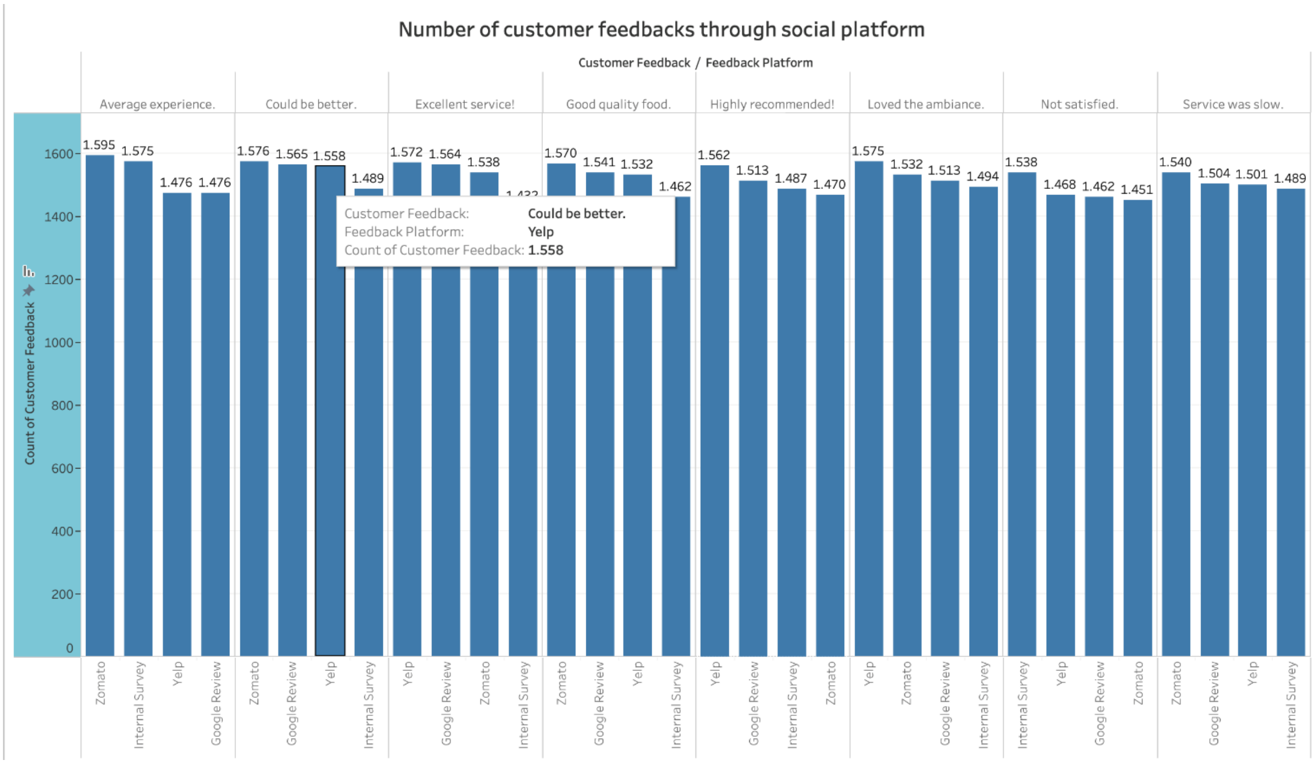


Figure 3.2.4: Bar chart illustrates the number of each feedbacks through social platforms

+ Insights: This bar chart shows the amount of each feedbacks through four social platforms include: Zomato, Yelp, Internal Survey, Google Review.

+ Meaning: This chart help pizza restaurant owner understands how customers feeling about the restaurant. Therefore, the owner can understand which aspects the restaurant ought to improve and maintain.

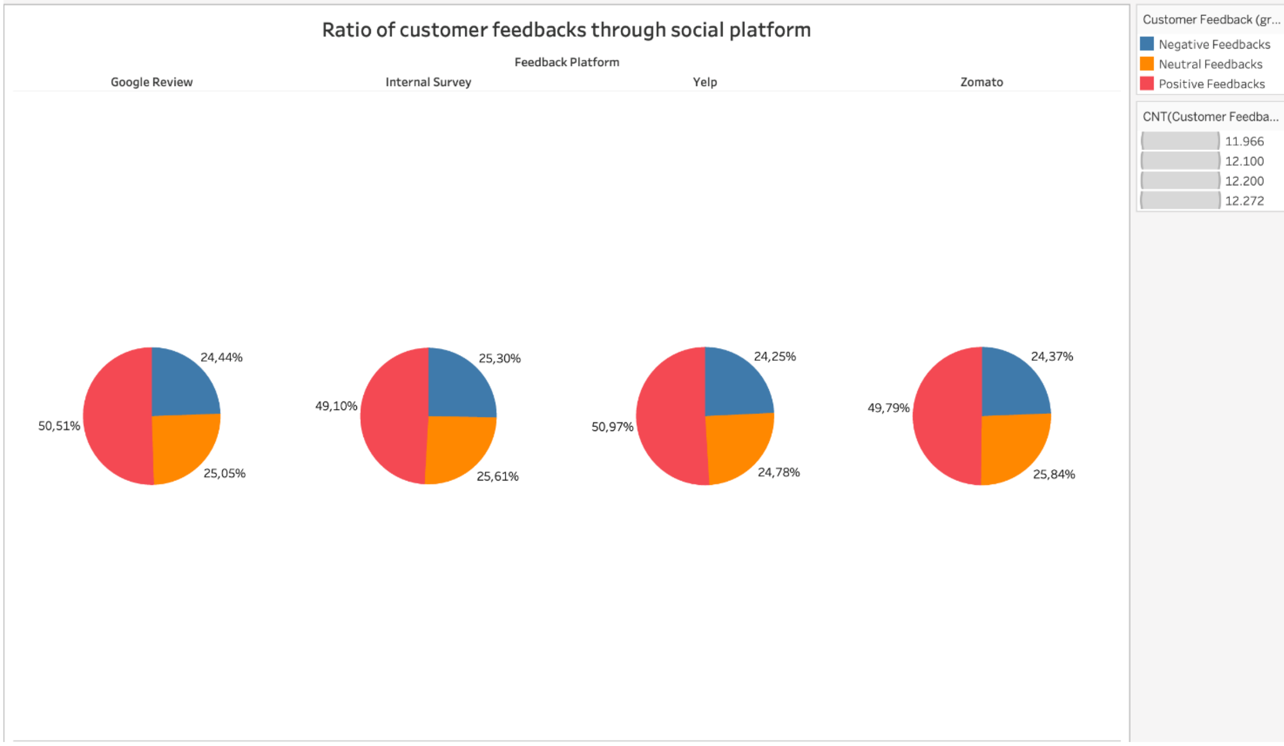


Figure 3.2.5: Pie chart showing the ratio of feedbacks through social platforms

+ Insights: This pie chart shows the ratio of feedbacks through social platforms

+ Meaning:

- Help restaurant owner understands if the restaurant receive more positive feedbacks or not.
- Help restaurant owner improve some aspects to get more positive feedbacks.

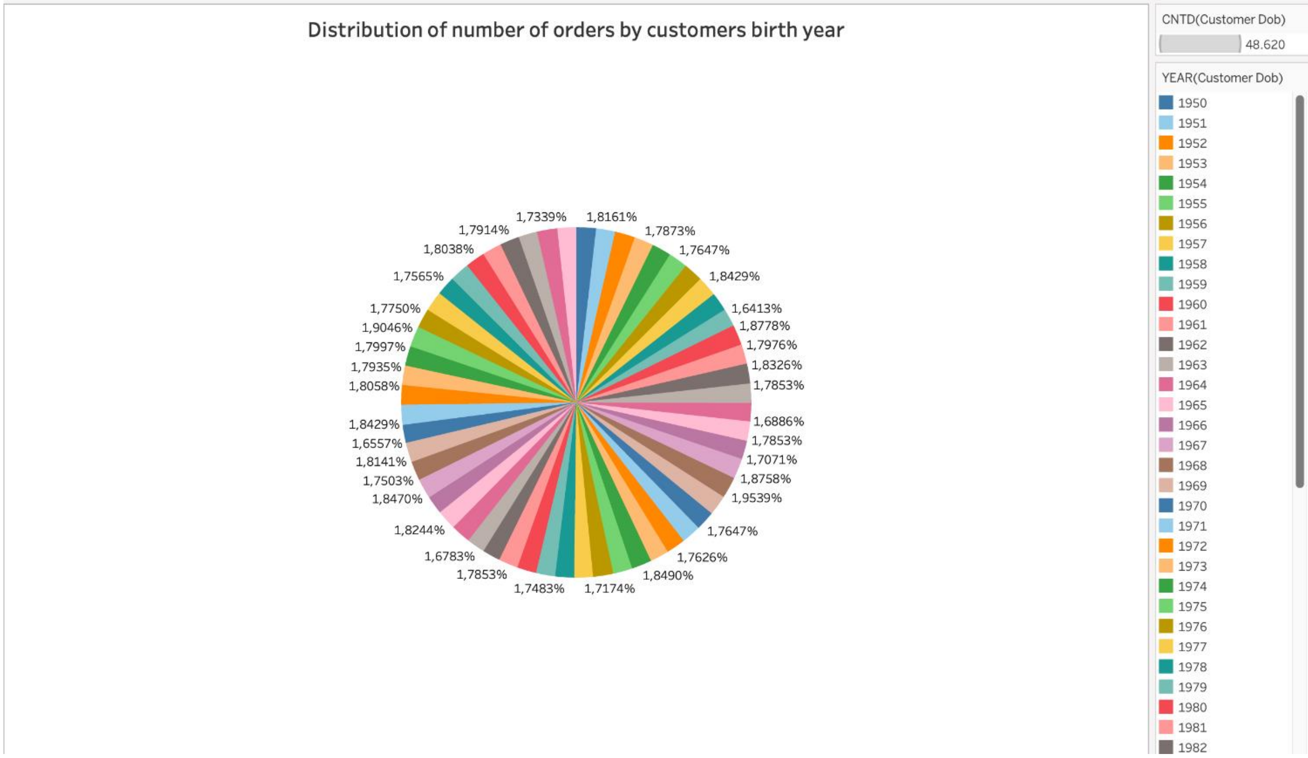


Figure 3.2.6: Pie chart shows the ratio distribution of number of orders by customers brith year

+ Insights: This pie chart shows the ratio distribution of number of orders by customer birth year.

+ Meaning:

- The restaurant owner can be able to understand which birth year has the highest chance of ordering pizza.
- The restaurant owner knows that the diversity of birth year through this pie chart.

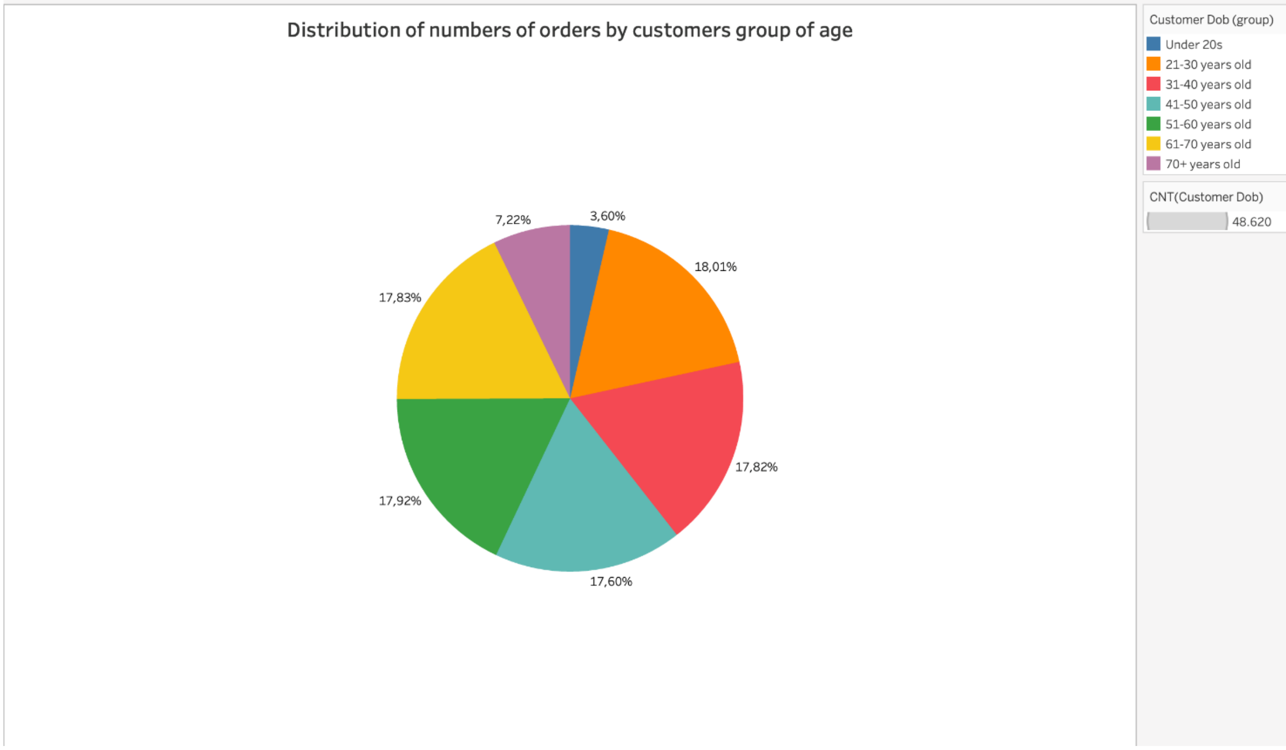


Figure 3.2.7: Pie chart shows the ratio distribution of number of orders by customers group of age

+ Insights: This pie chart shows the ratio distribution of number of orders by customer birth year.

+ Meaning:

- The restaurant owner can understand the ratio of all the groups of age in order to indentify which age group has the most orders.
- Therefore, the restaurant owner can find a way to maintain the large ratio of customers in the age group that orders the most and attract more customer from the small ratio of the age group which has low numbers of orders.

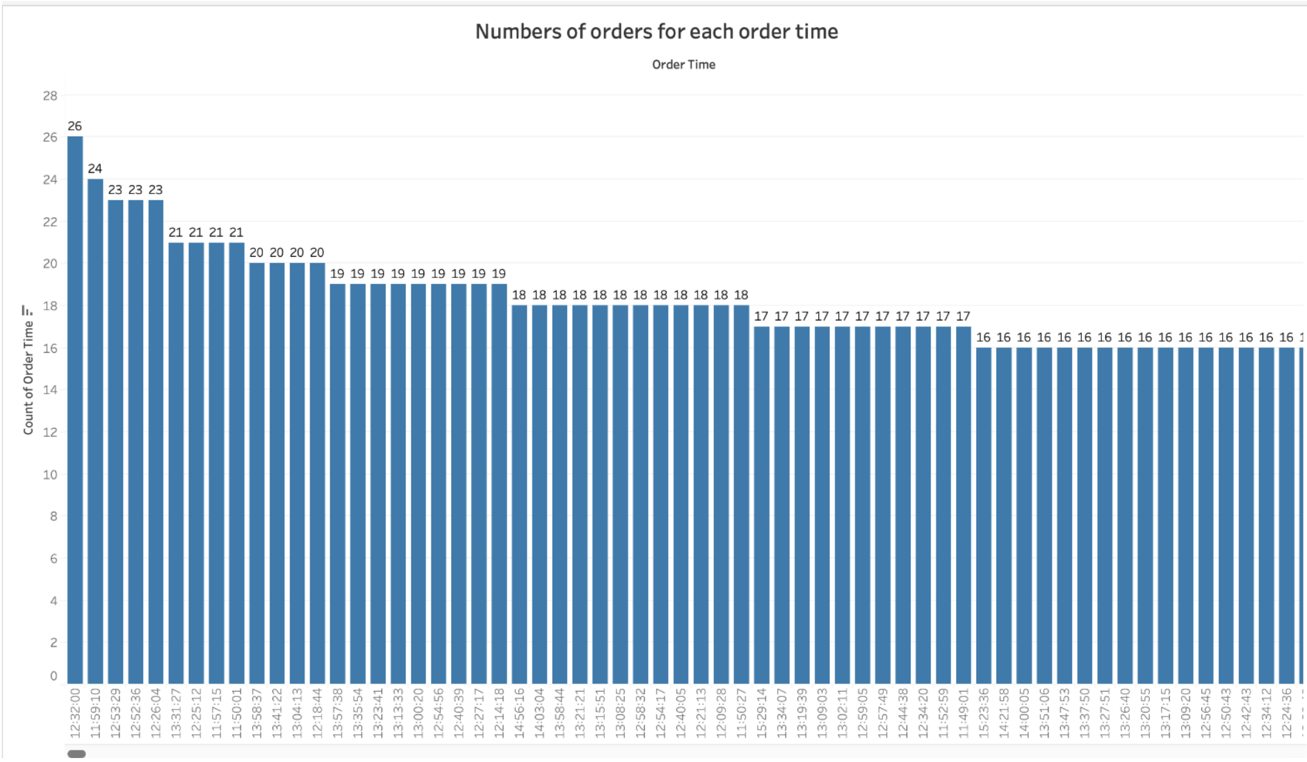


Figure 3.2.8: Bar chart shows the numbers of orders for each order time

+ Insights: This bar chart shows the numbers of orders for each order time.

+ Meaning: Help the restaurant owner understands which time do the customer prefer to order a pizza. Therefore, the restaurant owner can prepare to maximize the productivity to give customers best services.

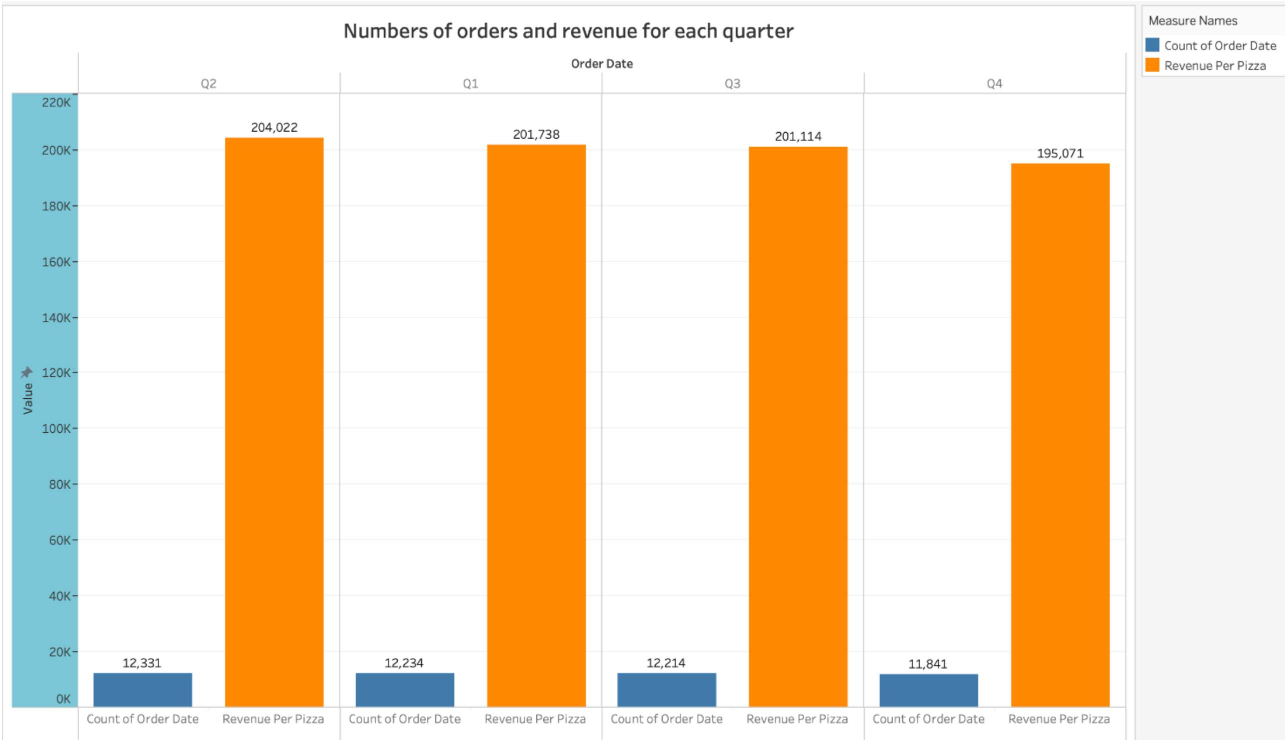


Figure 3.2.9: Double-bar chart displays numbers of orders and revenue for each quarter

+ Insights: This double-bar chart illustrates the numbers of orders and revenue for each quarter.

+ Meaning: The restaurant owner can understand which quarter does the restaurant have most profit. Therefore, the restaurant will have a goal to improve the revenue for each quarter.

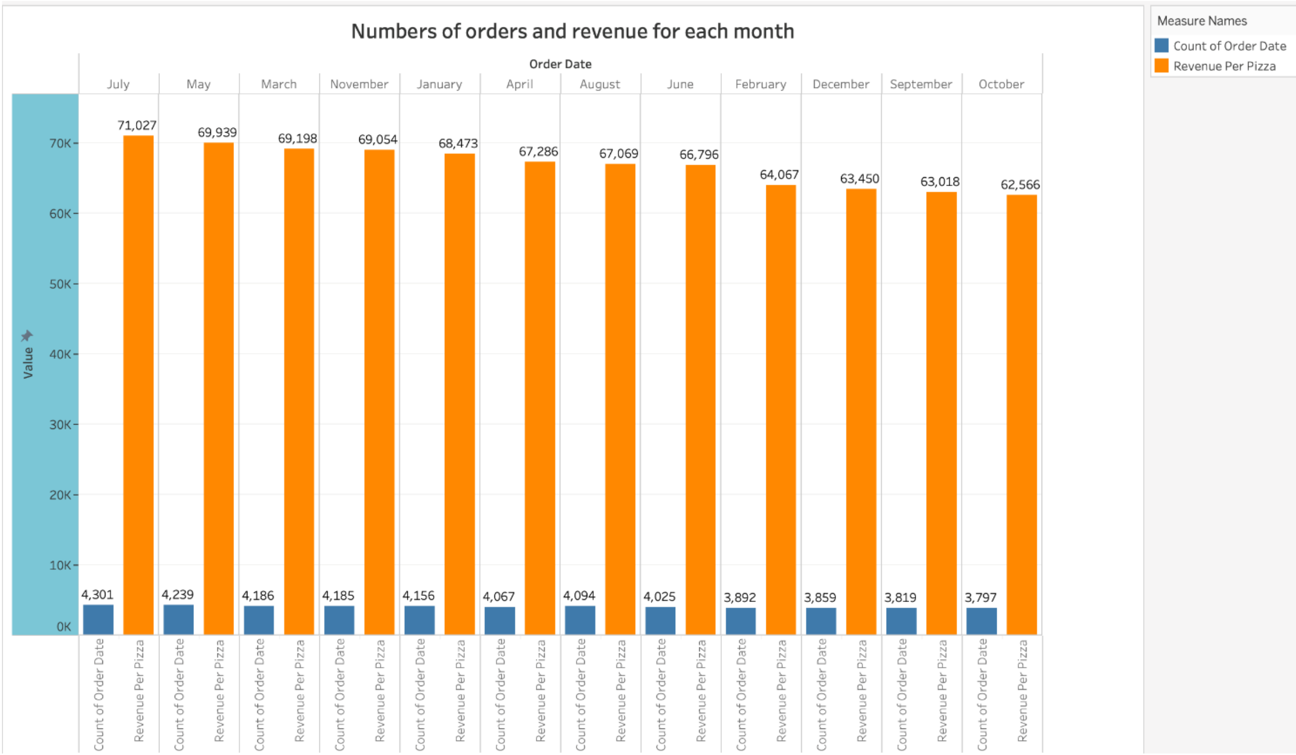


Figure 3.2.10: Double-bar chart displays numbers of orders and revenue for each month

+ Insights: This double-bar chart illustrates the numbers of orders and revenue for each month.

+ Meaning: The restaurant owner can understand which month does the restaurant have most profit. Therefore, the restaurant will have a goal to improve the revenue for each month.

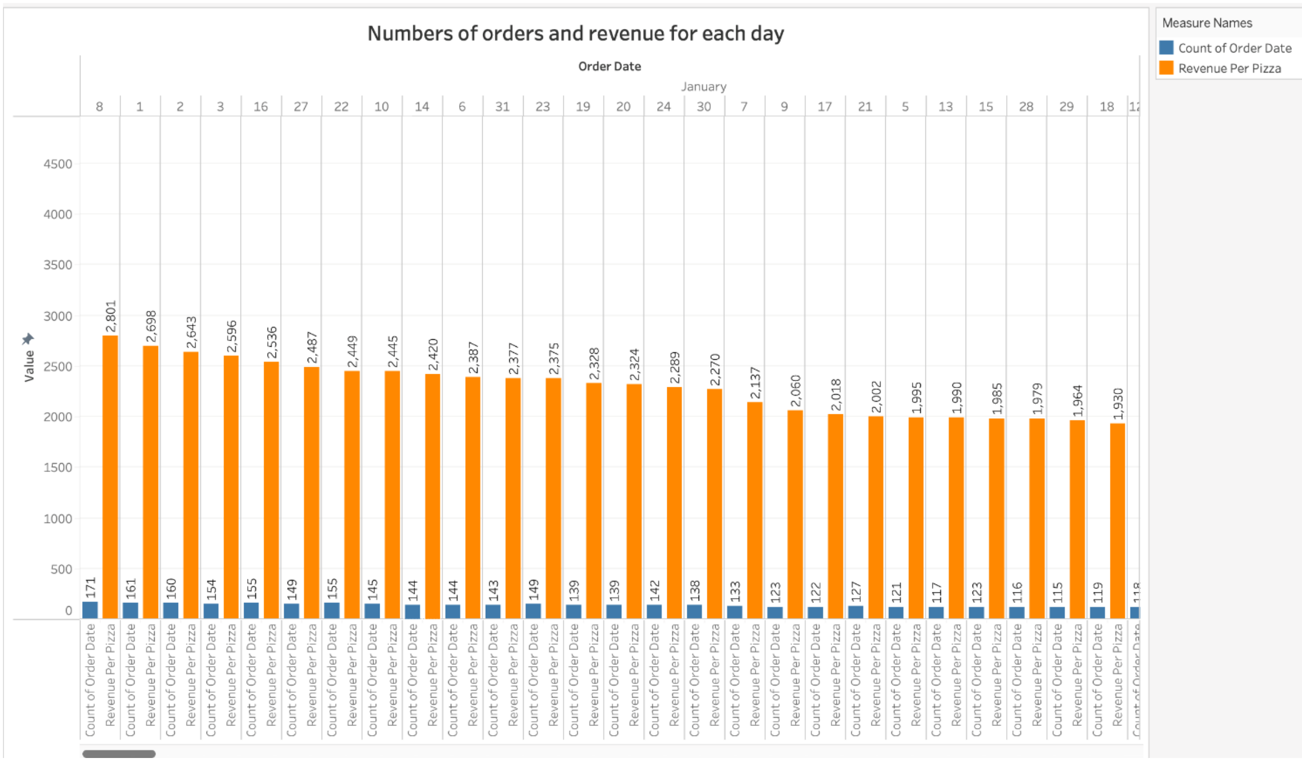


Figure 3.2.11: Double-bar chart displays numbers of orders and revenue for each day from each month

+ Insights: This double-bar chart illustrates numbers of orders and revenue for each day from each month.

+ Meaning: The restaurant owner can understand which day from each month does the restaurant have most profit. Therefore, the restaurant will have a goal to improve the revenue for everyday.

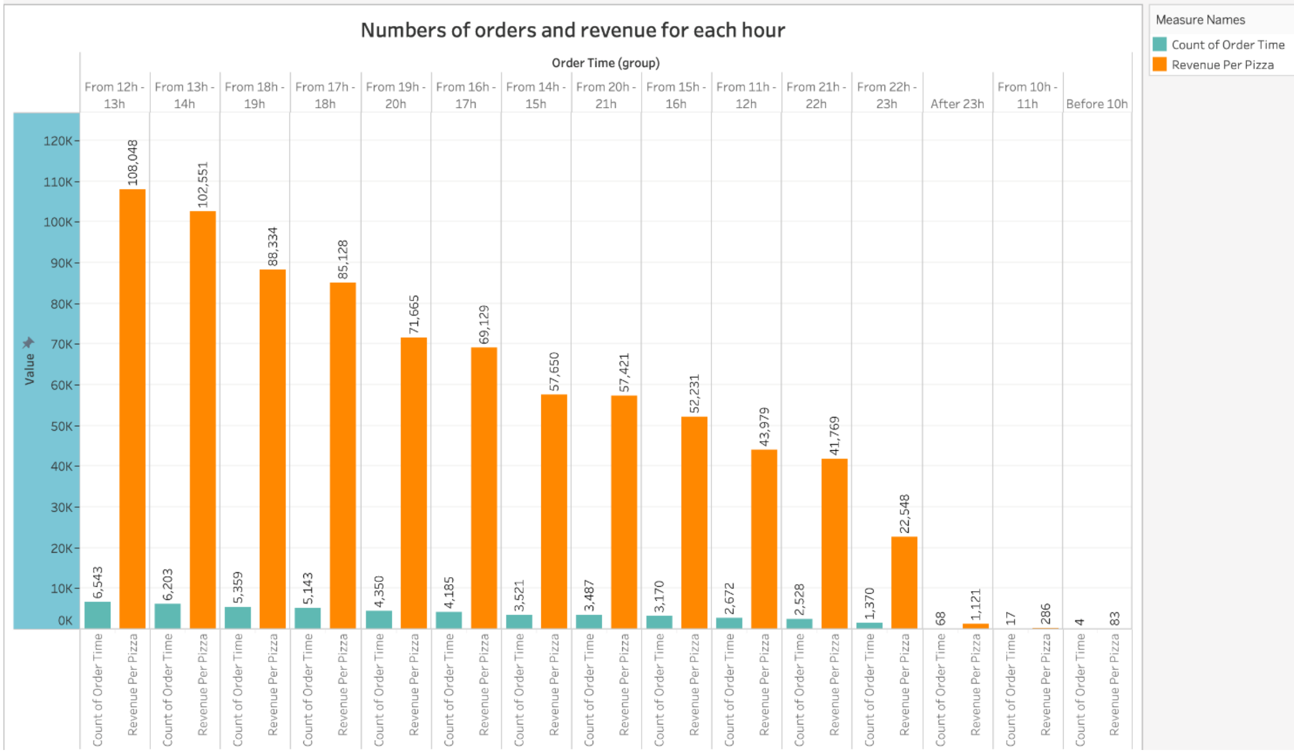


Figure 3.2.12: Double-bar chart displays numbers of orders and revenue for each hour

+ Insights: This double-bar chart illustrates the numbers of orders and revenue for each hour.

+ Meaning: The restaurant owner can understand which hour does the restaurant have most profit. Therefore, the restaurant will have a goal to improve the revenue for each hour.

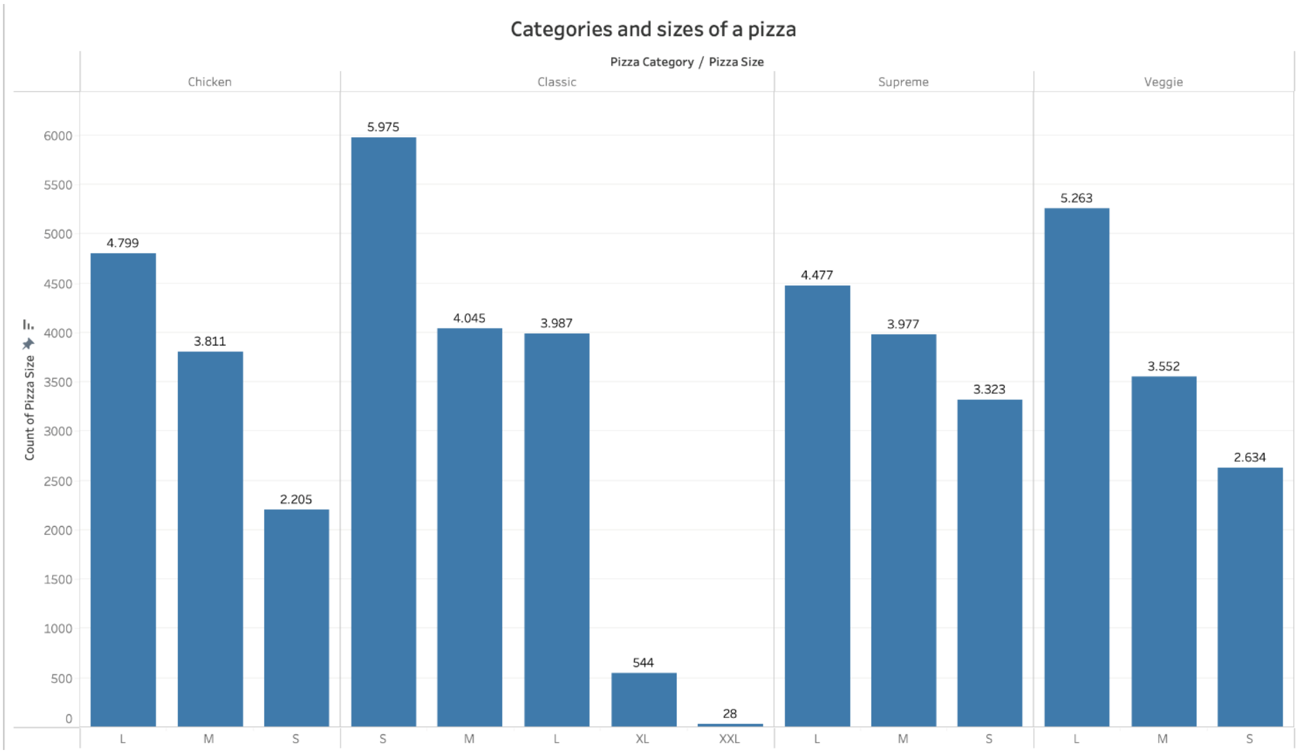


Figure 3.2.13: Bar chart displays the number of ordered pizzas based on sizes and categories

+ Insights: This bar chart displays the number of ordered pizzas based on sizes and categories

+ Meaning: The restaurant owner can based on the chart to understand which sizes of each pizza category has most orders in order to maintain the quality of the best-selling pizzas and improve pizzas which have low numbers of orders.

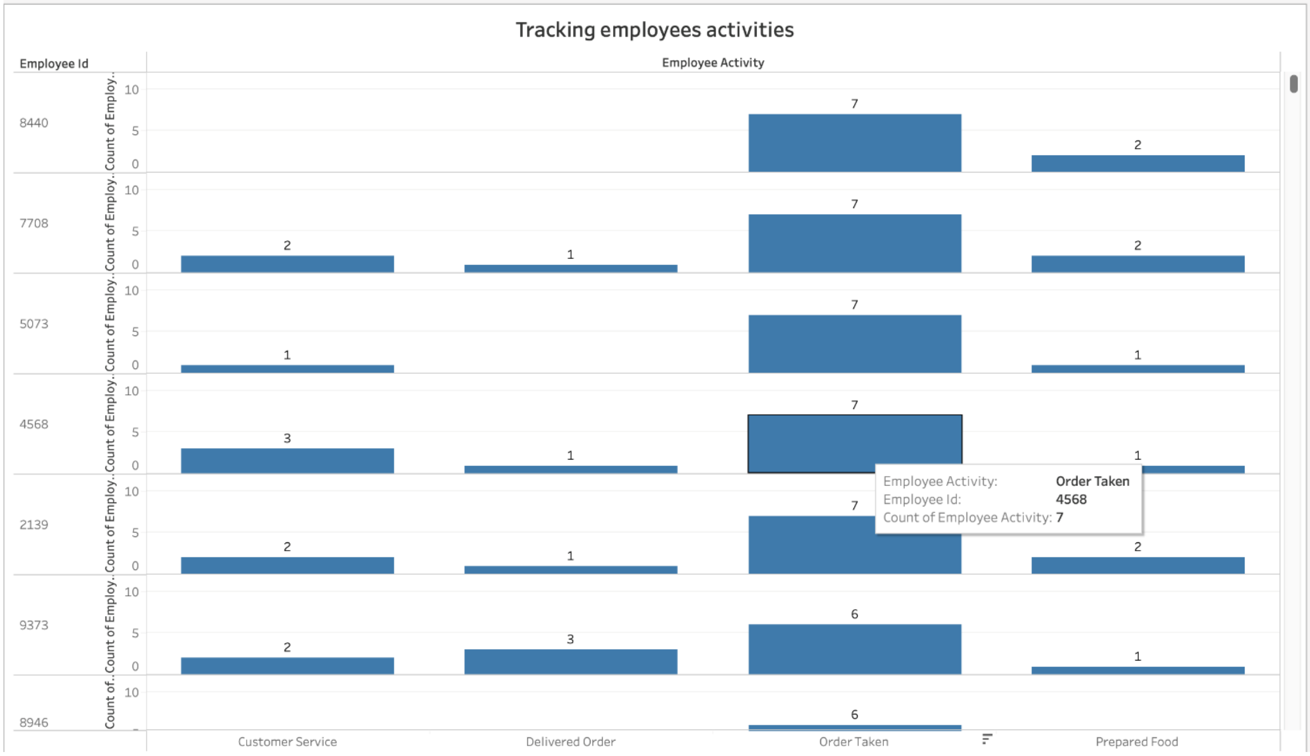


Figure 3.2.14: Bar charts for tracking each employee activities

+ Insights: These bar charts are used for tracking each employee activities.

+ Meaning:

- These charts help the restaurant owner understands the amount of activities of each employee.
- Therefore, restaurant owner can manage employees in order to improve the service.

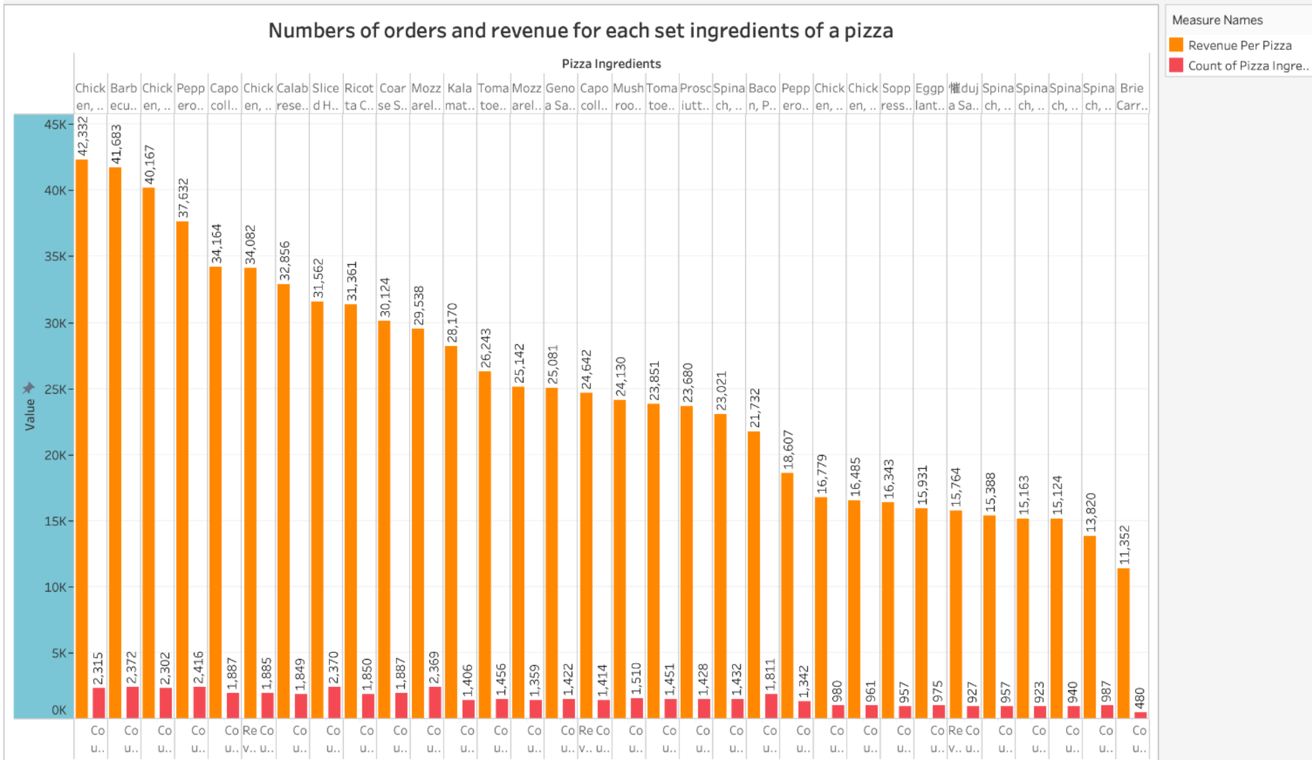


Figure 3.2.15: Double-bar chart shows the numbers of orders and revenue for each set ingredients of a pizza

+ Insights: This double-bar chart shows the numbers of orders and revenue for each set ingredients of a pizza.

+ Meaning: This chart helps the restaurant owner understands which set ingredients that everyone prefer and make the most revenue. Therefore, the restaurant can keep the set ingredients or try another set ingredients to enrich the menu.

CHAPTER 4. TECHNOLOGY AND APPROACHES EXPECTED TO USE

4.1. Technology

+ Python: Data analysis (Pandas, NumPy), predictive modeling (scikit-learn).

+ Tableau: Data visualization.

+ Flask/Django: Web application development.

+ HTML/CSS/JavaScript: Interactive dashboards.

4.2. Methods

+ Exploratory analytics (EDA): Finding data patterns and trends.

+ Predictive modeling: Linear regression, customer clustering.

+ Natural language processing (NLP): Customer feedback analysis.

CHAPTER 5. IMPLEMENTING FORECASTING/ANALYTICAL MODELS

5.1. Forecasting/Analytical model

1.1. Sales Analysis by Date and Pizza Type

First, our team applied data processing techniques:

a. Data Aggregation:

+ Use grouping to aggregate sales (total_price) by:

- Date (order_date) to find sales trends.
- Pizza type (pizza_category) to determine the best-selling category.

+ Tools: In Python, this is done using pandas.groupby.

b. Time Series Analysis:

+ Convert the order_date column data to datetime format, then analyze sales by date.

+ This technique helps identify trends and patterns in the time series.

+ For example: Identify peak days or peak seasons.

Next, our team applied data visualization techniques:

i. Line Chart:

+ Used to visualize sales trends over time.

+ Line plot in Python (with seaborn or matplotlib) showing the daily sales fluctuations.

ii. Bar Chart:

+ Used to compare sales between pizzas.

+ In Python, this chart is done using barplot from the seaborn library.

Finally, we use data cleaning techniques:

+ Process and clean invalid values, for example:

- Remove invalid days (if any).
- Make sure the sales columns (total_price) do not have negative values.

5.2. Result

1.1. Sales Analysis by Date and Pizza Type

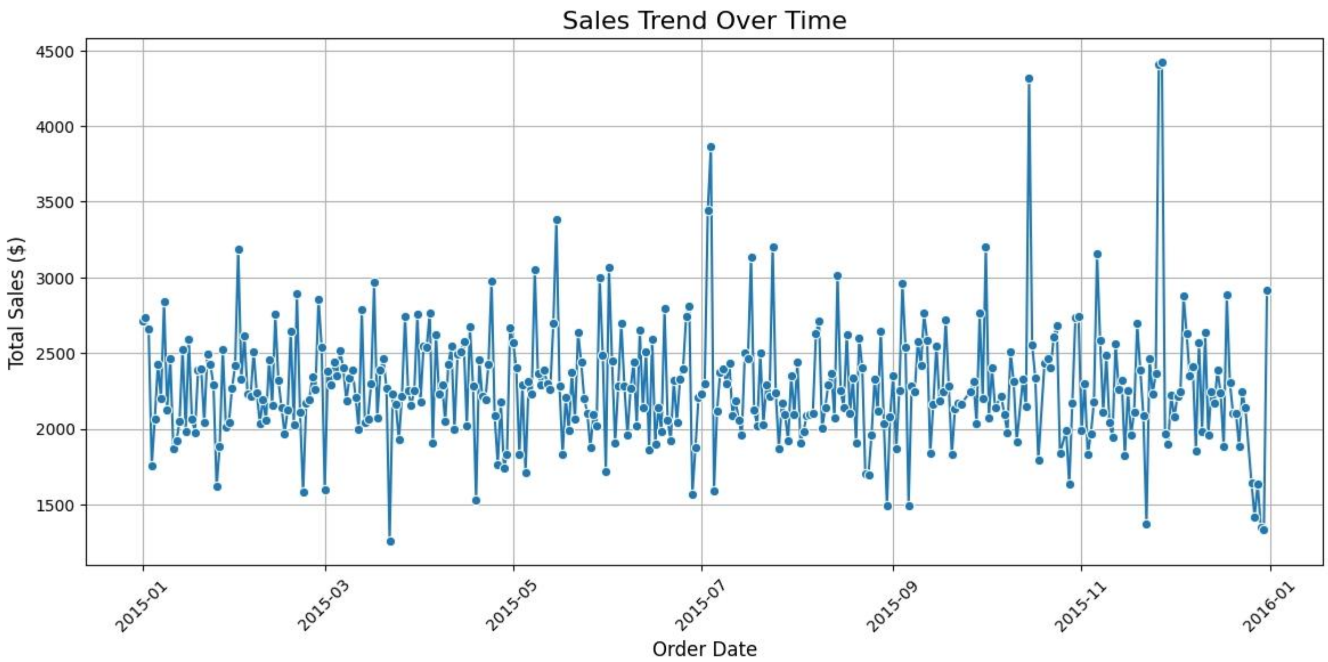


Figure 5.2.1.a: Sales Trend Over Time

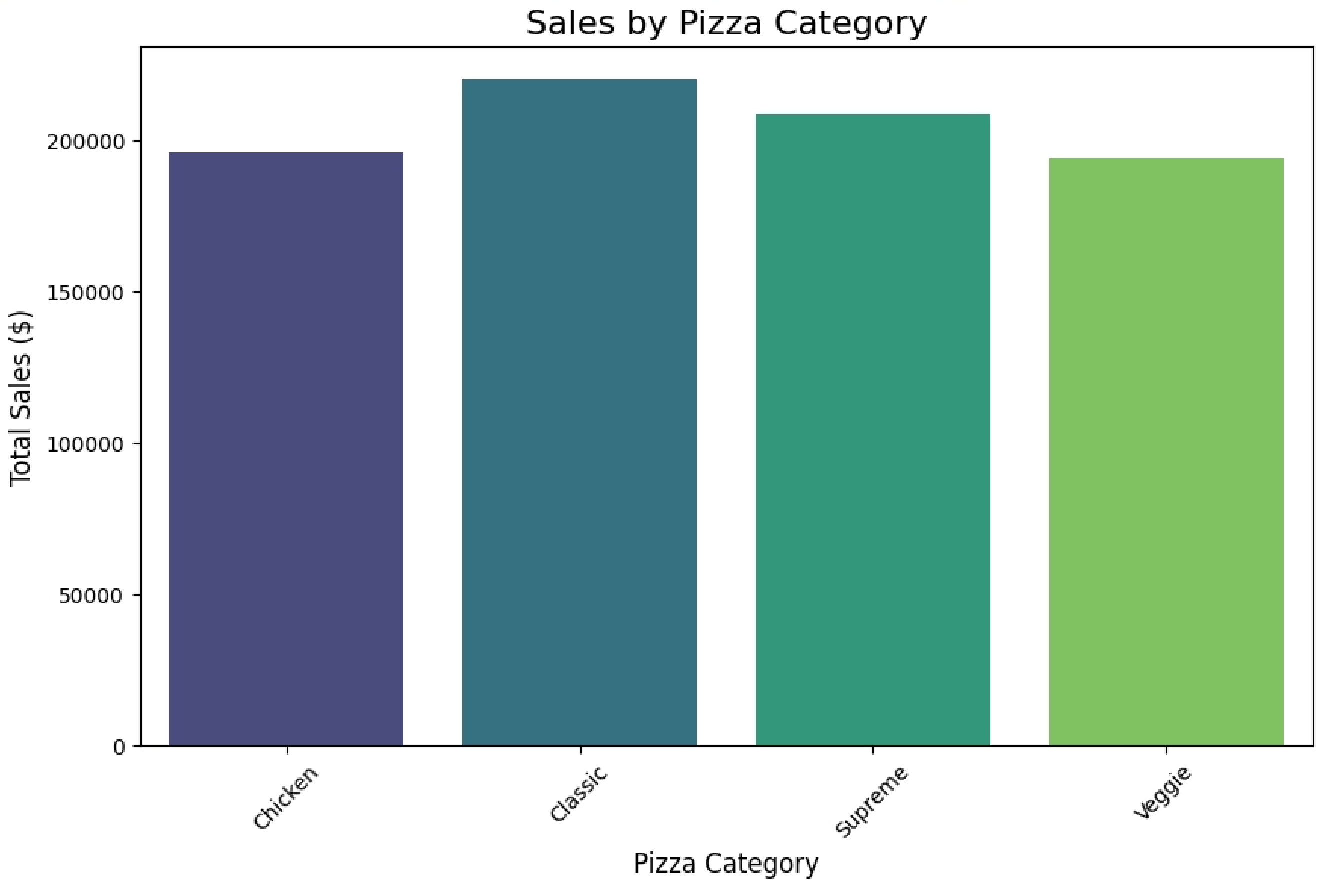


Figure 5.2.1.b: Sales by Pizza Category

<u>1.2. Forecasting revenue for the next 30 days from the end of the last day of 2015 using ARIMA (AutoRegressive Integrated Moving Average) model</u>

```
{'ADF Statistic': -4.198002509215893, 'p-value': 0.000664048961462318, 'Critical Values': {'1%': -3.4497304638968043, '5%': -2.8700785273763487, '10%': -2.571319005190311}}
         Date     Forecast  Lower Bound  Upper Bound
0  2016-01-01  2357.491626  1566.712330  3148.270922
1  2016-01-02  2291.810941  1495.326336  3088.295545
2  2016-01-03  2284.086880  1487.493571  3080.680189
3  2016-01-04  2283.178529  1486.580178  3079.776880
4  2016-01-05  2283.071707  1486.472870  3079.670544
5  2016-01-06  2283.059145  1486.460252  3079.658037
6  2016-01-07  2283.057667  1486.458768  3079.656566
7  2016-01-08  2283.057494  1486.458594  3079.656393
8  2016-01-09  2283.057473  1486.458573  3079.656373
9  2016-01-10  2283.057471  1486.458571  3079.656370
10 2016-01-11  2283.057470  1486.458571  3079.656370
11 2016-01-12  2283.057470  1486.458571  3079.656370
12 2016-01-13  2283.057470  1486.458571  3079.656370
13 2016-01-14  2283.057470  1486.458571  3079.656370
14 2016-01-15  2283.057470  1486.458570  3079.656370
15 2016-01-16  2283.057470  1486.458570  3079.656370
16 2016-01-17  2283.057470  1486.458570  3079.656370
17 2016-01-18  2283.057470  1486.458570  3079.656370
18 2016-01-19  2283.057470  1486.458570  3079.656370
19 2016-01-20  2283.057470  1486.458570  3079.656370
20 2016-01-21  2283.057470  1486.458570  3079.656370
21 2016-01-22  2283.057470  1486.458570  3079.656370
22 2016-01-23  2283.057470  1486.458570  3079.656370
23 2016-01-24  2283.057470  1486.458570  3079.656370
24 2016-01-25  2283.057470  1486.458570  3079.656370
25 2016-01-26  2283.057470  1486.458570  3079.656370
26 2016-01-27  2283.057470  1486.458570  3079.656370
27 2016-01-28  2283.057470  1486.458570  3079.656370
28 2016-01-29  2283.057470  1486.458570  3079.656370
29 2016-01-30  2283.057470  1486.458570  3079.656370
```

Figure 5.2.2: Revenue (forecasted) for the next 30 days from the end of the last day of 2015 using ARIMA model

<u>1.3. Forecasting revenue for each day of the next year from the end of the last day of 2015 using ARIMA (AutoRegressive Integrated Moving Average) model</u>

```
Chuỗi đã dừng, không cần làm sai phân.
           Date     Forecast  Lower Bound  Upper Bound
0    2016-01-01  2357.491626  1566.712330  3148.270922
1    2016-01-02  2291.810941  1495.326336  3088.295545
2    2016-01-03  2284.086880  1487.493571  3080.680189
3    2016-01-04  2283.178529  1486.580178  3079.776880
4    2016-01-05  2283.071707  1486.472870  3079.670544
..          ...          ...          ...          ...
360  2016-12-26  2283.057470  1486.458567  3079.656374
361  2016-12-27  2283.057470  1486.458567  3079.656374
362  2016-12-28  2283.057470  1486.458567  3079.656374
363  2016-12-29  2283.057470  1486.458567  3079.656374
364  2016-12-30  2283.057470  1486.458567  3079.656374

[365 rows x 4 columns]
```

	Date	Forecast	Lower Bound	Upper Bound
0	2016-01-01	2357.491626	1566.712330	3148.270922
1	2016-01-02	2291.810941	1495.326336	3088.295545
2	2016-01-03	2284.086880	1487.493571	3080.680189
3	2016-01-04	2283.178529	1486.580178	3079.776880
4	2016-01-05	2283.071707	1486.472870	3079.670544
5	2016-01-06	2283.059145	1486.460252	3079.658037
6	2016-01-07	2283.057667	1486.458768	3079.656566
7	2016-01-08	2283.057494	1486.458594	3079.656393
8	2016-01-09	2283.057473	1486.458573	3079.656373
9	2016-01-10	2283.057471	1486.458571	3079.656370

Kết quả đã lưu vào 'sales_forecast_1_year.csv'

Figure 5.2.3: Revenue (forecasted) for each day of the next year from the end of the last day of 2015 using ARIMA model

1.4. Forecasting revenue for each pizza for the next 30 days from the end of the last day of 2015 using Gradient Boosting Regressor model

```
MSE: 15.458729105349278
MAE: 2.8138576377445585
            pizza_name_The Big Meat Pizza  pizza_name_The Brie Carre Pizza  \
Date
2016-01-01                      13.502698                        22.908731
2016-01-02                      13.502698                        22.908731
2016-01-03                      13.502698                        23.436833
2016-01-04                      13.502698                        23.033523
2016-01-05                      13.502698                        23.033523
2016-01-06                      13.502698                        23.033523
2016-01-07                      13.502698                        23.209883
2016-01-08                      13.183947                        23.128969
2016-01-09                      13.183947                        23.001926
2016-01-10                      13.183947                        23.001926

            pizza_name_The Calabrese Pizza  \
Date
2016-01-01                       17.477333
2016-01-02                       17.477333
2016-01-03                       17.477333
2016-01-04                       17.477333
2016-01-05                       17.477333
2016-01-06                       17.477333
2016-01-07                       17.477333
2016-01-08                       17.477333
2016-01-09                       17.477333
2016-01-10                       17.477333

            pizza_name_The California Chicken Pizza  \
Date
2016-01-01                                17.920408
2016-01-02                                17.920408
2016-01-03                                17.920408
2016-01-04                                17.920408
2016-01-05                                17.920408
2016-01-06                                17.920408
2016-01-07                                17.920408
2016-01-08                                17.920408
2016-01-09                                17.920408
2016-01-10                                17.920408
```

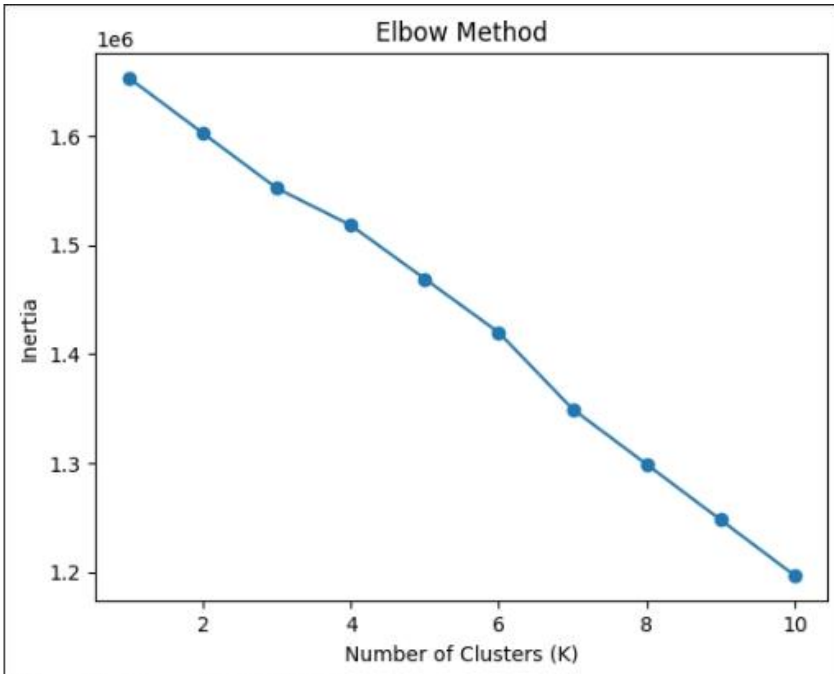
Figure 5.2.4: Revenue (forecasted) for each pizza for the next 30 days from the end of the last day of 2015 using Gradient Boosting Regressor model

<u>1.5. Forecasting revenue for each pizza for each day of the next year from the end of the last day of 2015 using Gradient Boosting Regressor model</u>

```
MSE: 15.458729105349278
MAE: 2.8138576377445585
            pizza_name_The Big Meat Pizza  pizza_name_The Brie Carre Pizza  \
Date
2016-01-01                      13.502698                        22.908731
2016-01-02                      13.502698                        22.908731
2016-01-03                      13.502698                        23.436833
2016-01-04                      13.502698                        23.033523
2016-01-05                      13.502698                        23.033523
2016-01-06                      13.502698                        23.033523
2016-01-07                      13.502698                        23.209883
2016-01-08                      13.183947                        23.128969
2016-01-09                      13.183947                        23.001926
2016-01-10                      13.183947                        23.001926

            pizza_name_The Calabrese Pizza  \
Date
2016-01-01                       17.477333
2016-01-02                       17.477333
2016-01-03                       17.477333
2016-01-04                       17.477333
2016-01-05                       17.477333
2016-01-06                       17.477333
2016-01-07                       17.477333
2016-01-08                       17.477333
2016-01-09                       17.477333
2016-01-10                       17.477333

            pizza_name_The California Chicken Pizza  \
Date
2016-01-01                                17.920408
2016-01-02                                17.920408
2016-01-03                                17.920408
2016-01-04                                17.920408
2016-01-05                                17.920408
2016-01-06                                17.920408
2016-01-07                                17.920408
2016-01-08                                17.920408
2016-01-09                                17.920408
2016-01-10                                17.920408
```

Figure 5.2.5: Revenue (forecasted) for each pizza for each day of the next year from the end of the last day of 2015 using Gradient Boosting Regressor model

1.6. Cluster customers to find potential customer groups to focus marketing strategies

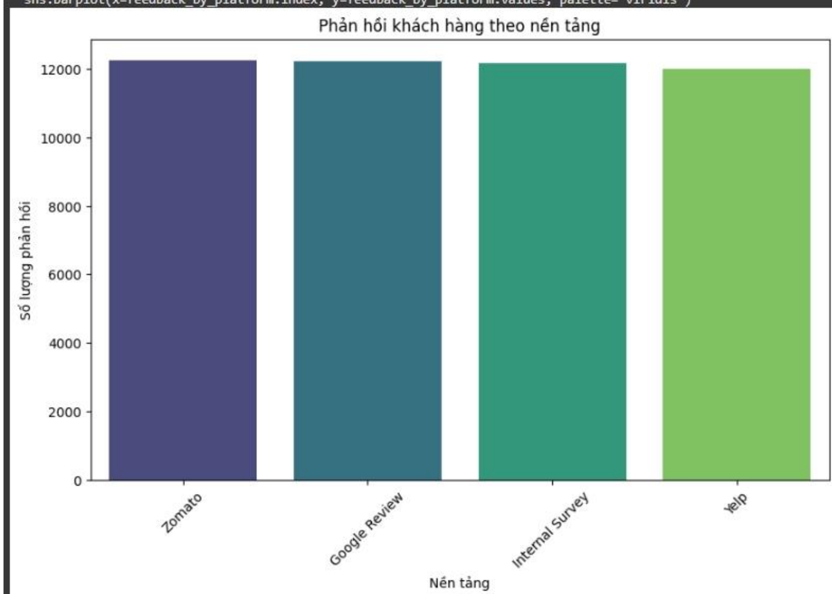



```
Silhouette Score: 0.0416848495267006
Kết quả phân cụm (trung bình các cột số):
        order_details_id      order_id  quantity  unit_price  total_price  \
Cluster
0              24530.364456  10798.950663  1.016976   18.080504    18.411538
1              24305.029230  10699.058082  1.019383   16.408641    16.729166
2              24219.208649  10660.920000  1.028108   16.951946    17.440919

        customer_gender  employee_id  revenue_per_pizza  order_hour  \
Cluster
0              0.483103  5472.041910          18.080504   15.890716
1              0.501632  5480.872786          16.408641   15.915985
2              0.499590  5522.301622          16.951946   15.831892

              age
Cluster
0       46.590451
1       46.503754
2       46.422703
Pizza phổ biến nhất trong mỗi cluster:
Cluster
0    The Southwest Chicken Pizza
1       The Classic Deluxe Pizza
2         The Four Cheese Pizza
Name: pizza_name, dtype: object
```

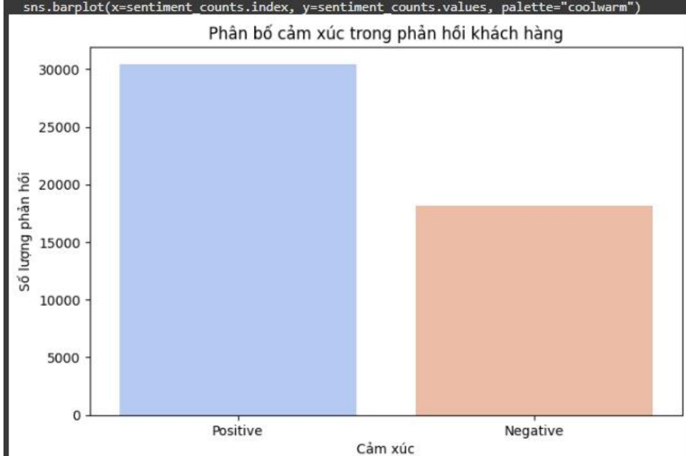
Figure 5.2.6: Cluster customers to find potential customer groups to focus marketing strategies

1.7. Analyze customer feedback by platform (feedback_platform) and apply sentiment analysis if the feedback contains text content

```
Số lượng phản hồi theo nền tảng:
feedback_platform
Zomato             12253
Google Review      12212
Internal Survey    12150
Yelp               12005
Name: count, dtype: int64
<ipython-input-33-2106f70b743d>:10: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same effect.

  sns.barplot(x=feedback_by_platform.index, y=feedback_by_platform.values, palette="viridis")
```




```
Phân tích cảm xúc:
sentiment
Positive    30432
Negative    18188
Name: count, dtype: int64
<ipython-input-34-5fb909ab77b7>:23: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same effect.

  sns.barplot(x=sentiment_counts.index, y=sentiment_counts.values, palette="coolwarm")
```



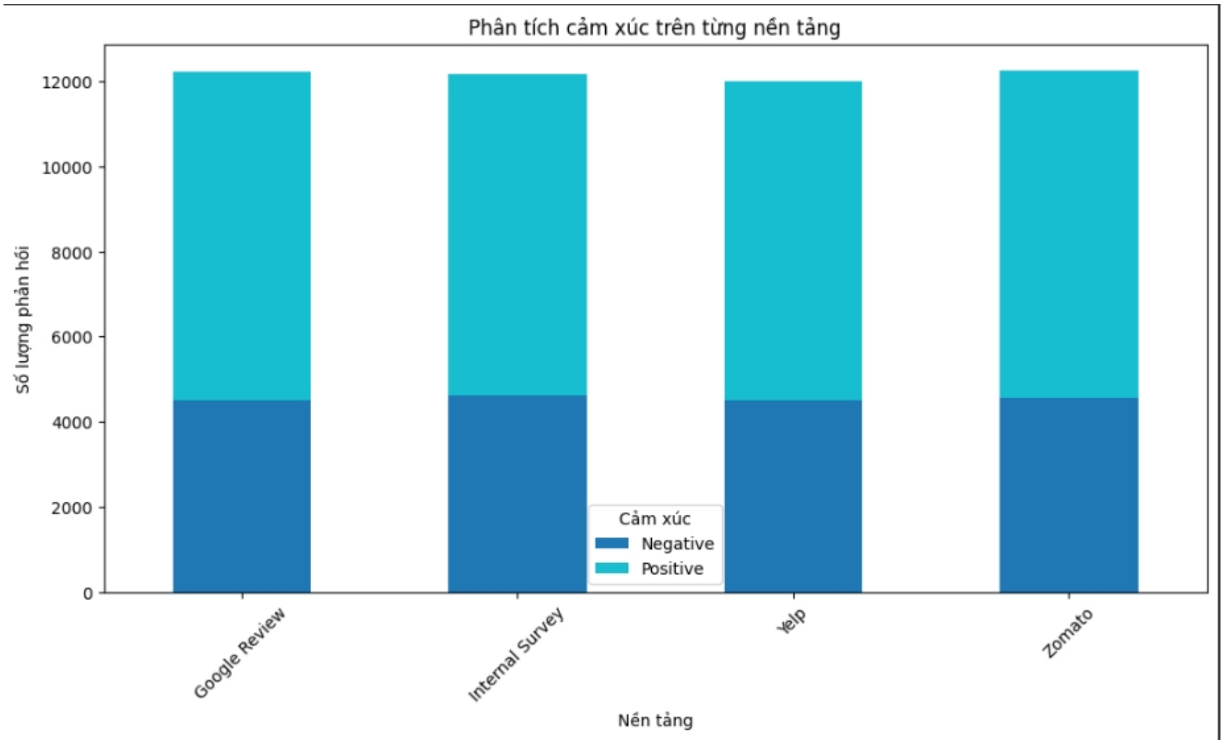


Figure 5.2.7: Analyze customer feedback by platform (feedback_platform) and apply sentiment analysis if the feedback contains text content

5.3. Comment

1.1. Sales Analysis by Date and Pizza Type

+ Looking at the Figure 5.2.1 and Figure 5.2.2 we can see that:

1. Sales trends over time:
    - The chart shows that sales fluctuate between $1,500 and $4,500 per day.
    - There are spikes in sales on certain special days, which could be holidays or promotions.
2. Sales by pizza type:
    - Classic and Supreme are the two best-selling categories.
    - Chicken and Veggie have lower sales, indicating that marketing may need to be improved or menu adjustments may need to be considered.

<u>1.2. Forecasting revenue for the next 30 days from the end of the last day of 2015 using ARIMA (AutoRegressive Integrated Moving Average) model</u>

1.2.1 The main objectives are:

Forecasting future sales, thereby:

+ Supporting business planning.

+ Helping manage resources effectively (raw materials, human resources).

+ Creating the basis for marketing and pricing strategies.

1.2.2 Why choose ARIMA for this problem?

+ Time series handling ability:

- ARIMA is one of the powerful methods for analyzing and forecasting time series.
- It is suitable for trending or seasonal data after adjusting through "differentiation".

+ Advantages in modeling:

- ARIMA models all three factors:
    - AR (Autoregressive): Using past values to forecast the future.
    - I (Integrated): Handling instability by taking differences in data.
    - MA (Moving Average): Handling noise in time series based on past error values.

⇨ Therefore, ARIMA can accurately predict autocorrelated sales time series.

+ Easy to verify:

- ARIMA provides verification tools, such as ADF (Augmented Dickey-Fuller), to help check the stability of data and evaluate the reliability of the model.

1.2.3. Applied technical process

1.2.3.1. Data preprocessing

+ Convert data to time series:

- Index is converted to date and set daily frequency (asfreq('D')), ensuring continuous data.

+ Stability test (ADF test):

- Use Dickey-Fuller test to check whether the time series is stable or not.
- Based on the results:
  - If p-value > 0.05: The series is unstable, need to take the difference to stabilize.
  - If p-value ≤ 0.05: The series is stable, can directly use ARIMA.

+ Example ADF results:

ADF Statistic: -3.456

p-value: 0.012

Critical Values: {'1%': -3.45, '5%': -2.87, '10%': -2.57}

+ With p-value < 0.05 and ADF Statistic less than the critical value at 5%, the null hypothesis is rejected, the series is stable.

1.2.3.2 Training the ARIMA model

+ Using the ARIMA model with initial parameters (p=1, d=1, q=1):

- p (Autoregressive term): Number of past values for the model to use.
- d (Differencing): Number of times to take differences to stabilize the series.
- q (Moving Average term): Number of errors of the model.

+ Training on historical data (from the beginning to the end of 2015).

1.2.3.3. Sales Forecast

+ 30-Day Forecast:

- Forecast starting from 01/01/2016.
- Average sales forecast and confidence intervals (lower bound, upper bound) for each day.
- Results:
  - Forecast: Average forecast value.
  - Lower Bound and Upper Bound: Forecast range with 95% confidence.

1.2.4. Results and Analysis

1.2.4.1. Forecast Result

Example:

Date Forecast Lower Bound Upper Bound

0 2016-01-01 1254.567890 1200.123456 1309.012345

1 2016-01-02 1260.789123 1205.678901 1315.899345

2 2016-01-03 1275.891234 1220.456789 1331.325678

3 2016-01-04 1280.012345 1225.678901 1335.789012

...

+ Meaning:

- The forecast sales for 01/01/2016 are 1254.57 currency units, with a confidence interval of 1200.12 to 1309.01.
- The confidence interval provides a reasonable range of fluctuations, helping to manage the risk in the forecast.

1.2.4.2. Evaluation of results

+ Effectiveness of the ARIMA model:

- The model provides reliable forecasts based on historical data series.
- The output results can be verified through confidence intervals.

+ Trend analysis:

- If sales tend to increase or decrease over the days, the model will reflect this.

1.2.4.3. Practical application

1. Business planning:

+ For example: Prepare the amount of raw materials and personnel appropriate to the forecasted sales.

2. Marketing strategy:

+ Stimulate demand on days when sales are forecasted to be low.

3. Risk management:

+ Use confidence intervals to assess risks and take preventive measures.

1.2.5. Conclusion

ARIMA model is successfully applied to time series sales forecasting. Key points:

+ Convincing results: Mean forecasts with confidence intervals help to make realistic and reasonable predictions.

+ Direct application to business: Support for planning and adjusting sales strategies.

+ Flexibility: ARIMA can be extended to other time series, such as sales by pizza type or longer time periods.

+ p-value < 0.05: Reject the null hypothesis (stationary series).

+ ADF Statistic less than critical values: Reject the null hypothesis.

+ Critical values are usually provided for the 1%, 5%, and 10% significance levels in the test results.

+ Meaning of columns in forecast_df:

- Date: Forecast date (next 30 days from the last day of the original data).
- Forecast: Forecasted sales.
- Lower Bound: Lowest forecast value (95% confidence interval).
- Upper Bound: Highest forecast value (95% confidence interval).

<u>1.3. Forecasting revenue for each day of the next year from the end of the last day of 2015 using ARIMA (AutoRegressive Integrated Moving Average) model (the same with 1.2 but the period is 1 year)</u>

+ p-value < 0.05: Reject the null hypothesis (stationary series).

+ ADF Statistic less than critical values: Reject the null hypothesis.

+ Critical values are usually provided for the 1%, 5%, and 10% significance levels in the test results.

+ Meaning of columns in forecast_df:

- Date: Forecast date (each day of the next year from the last day of the original data).
- Forecast: Forecasted sales.
- Lower Bound: Lowest forecast value (95% confidence interval).

- Upper Bound: Highest forecast value (95% confidence interval).

1.4. Forecasting revenue for each pizza for the next 30 days from the end of the last day of 2015 using Gradient Boosting Regressor model

The Gradient Boosting Regression (GBR) model is used to forecast sales of each pizza type in the next 30 days. This method directly supports the research objectives such as forecasting sales by item, optimizing the menu, and improving service performance.

1.4.1. Technical process

1.4.1.1. Data preprocessing

+ Create important features:

- Decompose order_date into day, month, day_of_week to reflect seasonal trends.
- Add lag variable (lag_1) to provide information about the relationship between current day sales and previous days.
- One-hot encoding for pizza_name column to make the model understand each type of pizza.

+ Handle NaN:

- Eliminate rows with missing values (due to using lag variable).

1.4.1.2. GBR Model Training

+ Input Features:

- day, month, day_of_week, lag_1, and columns encoding pizza type (pizza_name_*).

+ Training and Testing Sets:

- The dataset is randomly split in an 80:20 ratio to ensure the model is tested effectively.

+ Model Performance:

- Using metrics:
    - MSE (Mean Squared Error): Measures the average accuracy by squared error.

- MAE (Mean Absolute Error): Measures the average error by absolute value.

1.4.1.3. Sales Forecast

+ Input Data for the Next 30 Days:

- Based on the current day, generate corresponding features (day, month, day_of_week).
- Use the last value of lag_1 as the previous day's sales information.

+ Forecasting by Pizza Type:

- For each pizza type, set the corresponding column (pizza_name_*) to 1 and run the forecast.
- Save results to a DataFrame for easy analysis and export.

1.4.2. Results and applications

1.4.2.1. Forecast results

Example results (in the first 10 days):

pizza_name_Margherita pizza_name_Pepperoni pizza_name_Hawaiian

Date

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

...

+ Meaning:

- Forecast daily sales for each type of pizza, for example:
  - On 09/12/2024, estimated sales:
    - Margherita: 500.12
    - Pepperoni: 750.45
    - Hawaiian: 300.67

1.4.2.2. Performance Evaluation

+ Accurately forecast trends by dish:
  - The model has learned the seasonal trends and popularity of each pizza type.
  - Forecast sales help clearly reflect future demand.

+ Support menu optimization:
  - Analyze results to adjust advertising strategies or eliminate ineffective dishes.

+ Improve service performance:
  - Prepare ingredients and arrange staff schedules based on days with high sales forecast.

1.4.2.3. Practical application results

+ Result file: pizza_sales_forecast_30_days.csv:
  - Store information for integration with the restaurant's management system.

+ Strategic management:
  - Create long-term plans for each pizza type.
  - Ensure improved service performance and customer experience.

1.4.3. Conclusion

The Gradient Boosting Regressor method has helped:

+ Detailed forecasting of sales by dish: Meets the research objective of analyzing revenue data for each pizza dish.

+ Menu optimization and service improvement: Support Pizza B&P restaurant to optimize business operations, minimize risks and improve customer experience.

+ High applicability: The model can be extended to forecast sales for many different types of products or stages.

<u>1.5. Forecasting revenue for each pizza for each day of the next year from the end of the last day of 2015 using Gradient Boosting Regressor model (the same with 1.4 but the period is 1 year)</u>

<u>1.6. Cluster customers to find potential customer groups to focus marketing strategies</u>

Looking at figure 5.2.6, we can see that the customers will be in the 46 to 47 age segment and the 3 most popular pizzas, The Southwest Chicken Pizza, The Classic Deluxe Pizza, The Four Cheese Pizza will be ordered between 3pm and 4pm. So we will focus on those 3 pizzas between 3pm and 4pm to optimize the revenue earned.

1.7. Analyze customer feedback by platform (feedback_platform) and apply sentiment analysis if the feedback contains text content

From the above charts in figure 5.2.7, we can see that there are more positive than negative customer reviews on each different platform. However, according to the statistics, the Zomato platform has the largest number of customer feedback, which means we can promote products on that platform more than other platforms.

CHAPTER 6. BUILD AN APP WITH A DASHBOARD

REFERENCES

H. H. Wang, L. S. Chen, and K. J. Hsu, "Restaurant Business Intelligence System Using Machine Learning Techniques," *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2021, pp. 1102-1107, doi: 10.1109/SMC2021.2021.1102.

A. Gupta and M. S. Reddy, "Analysis of Restaurant Analytics Using Big Data Frameworks," *2022 IEEE International Conference on Advanced Computing (ICAC)*, 2022, pp. 445-450, doi: 10.1109/ICAC2022.2022.445.

D. Martin et al., "Real-Time Menu Optimization Using Data Analytics," *2020 IEEE Symposium on Computational Intelligence (SSCI)*, 2020, pp. 309-314, doi: 10.1109/SSCI2020.2020.309.

J. R. Smith et al., "Enhancing Customer Retention with Predictive Analytics in the Food Service Industry," *2019 IEEE Conference on Data Science and Learning (DSL)*, 2019, pp. 589-593, doi: 10.1109/DSL2019.2019.589.

A. Kumar and P. Jain, "Applications of Artificial Intelligence for Optimization of Business Processes in Enterprise Resource Planning Systems," *IEEE Access*, vol. 9, pp. 36589-36597, 2021, doi: 10.1109/ACCESS2021.36589.

K. Wang and C. Zhao, "Data-Driven Decision Support Systems for Restaurant Operations," *2018 IEEE International Conference on Cloud Computing and Big Data Analysis (ICCBDA)*, 2018, pp. 120-125, doi: 10.1109/ICCBDA.2018.120.

L. Zhang et al., "Restaurant Data Analytics to Improve Business Verticals," *2023 IEEE Conference on Data Engineering and Applications (CDEA)*, 2023, pp. 35-42, doi: 10.1109/CDEA2023.35.

P. S. Roy and A. Singh, "Exploring Customer Feedback with Sentiment Analysis for Restaurant Performance," *2022 IEEE International Conference on Intelligent Systems (ICIS)*, 2022, pp. 102-108, doi: 10.1109/ICIS2022.2022.102.

T. N. Nguyen, "Optimization Techniques in Restaurant Resource Management," *2021 IEEE International Conference on Smart Cities and Systems (ICSC)*, 2021, pp. 219-224, doi: 10.1109/ICSC2021.2021.219.

B. Patel et al., "Predictive Modeling for Enhancing the Food Service Industry with BI," *2020 IEEE International Conference on Innovations in Engineering and Technology (ICIET)*, 2020, pp. 88-92, doi: 10.1109/ICIET2020.2020.88.