Foodhub Data Analysis – Alan Mc Girr

Python Foundations

16-03-2025

Contents / Agenda

- Executive Summary
- Business Problem Overview and Solution Approach
- Data Overview
- EDA - Univariate Analysis
- EDA - Multivariate Analysis
- Appendix

Executive Summary

- Please mention conclusions, actionable insights & recommendations
- Please add the answer for question 17

***Note**: You can use more than one slide if needed*

Data Insights

1

Revenue Drivers

• High-value orders ($20+) generate the most revenue ($3,688.73), while low-value orders (≤$5) contribute $0 revenue due to the commission model.

• Shake Shack leads in revenue & order frequency, making it a strategic restaurant partner.

2

Delivery Time Patterns

• Weekday deliveries (28.34 min) are slower than weekends (22.47 min) due to traffic and demand variations.

• Japanese & American cuisines have the most delayed orders (60+ min), indicating a need for better order management.

3

Customer Ratings Analysis

• Delivery time does not significantly impact ratings (4.21 avg. for both fast & delayed orders).

• Top-rated restaurants (The Meatball Shop, Blue Ribbon Fried Chicken, Shake Shack) qualify for promotional offers.

Conclusions from Analysis

1

Revenue Drivers

Orders above $20 contribute most to revenue

2

Key Restaurant Partners

Shake Shack dominates in orders and revenue

3

Delivery Efficiency

Weekend deliveries are faster than weekday ones

- The analysis reveals that high-value orders ($20+) are the primary revenue drivers, contributing $3,688.73 compared to mid-value orders at $2,477.58.
- Shake Shack stands out as the top revenue-generating restaurant, dominating both in revenue and order volume.
- Weekday deliveries take longer (28.34 minutes) than weekend deliveries (22.47 minutes), suggesting different operational challenges.
- Customer ratings are not significantly impacted by delivery time, with both fast and slower deliveries receiving similar ratings.

Business Recommendations

Loyalty Program

Introduce a loyalty program to reward high-spending and repeat customers with exclusive discounts or free delivery on future orders.

Restaurant Partnerships

Work closely with high-revenue, high-rated restaurants like Shake Shack, The Meatball Shop, and Blue Ribbon to increase order fulfillment speed and maintain quality standards.

Dynamic Pricing

Implement dynamic delivery pricing for peak weekday hours, allowing customers to pay extra for faster deliveries when traffic and demand are high.

Delivery Transparency

Provide real-time tracking and estimated delivery time predictions based on cuisine type and restaurant efficiency to manage customer expectations.

Final Strategy Recommendations

1

Encourage Higher-Value Orders

Promote bundle deals, free delivery for orders over $20, or discount vouchers for higher spending to increase the proportion of high-value orders that generate 25% commission.

2

Feature Top-Rated Restaurants

Create special promotional campaigns for highly-rated restaurants like The Meatball Shop, Blue Ribbon Fried Chicken, and Shake Shack, such as "Highly Rated Favorites" or customer loyalty rewards.

3

Optimize Weekday Delivery

Address the longer weekday delivery times by implementing route optimization, increasing delivery staff during peak hours, or offering premium delivery options.

4

Focus on Food Quality

Since ratings are not strongly influenced by delivery time, work with restaurants to improve packaging, food presentation, and quality control measures to enhance overall customer satisfaction.

Top Restaurants for Promotional Offers

The Meatball Shop

Highest average rating of 4.33 among qualifying restaurants. Second-highest revenue generator at $2,145.21.

Blue Ribbon Fried Chicken

Strong average rating of 4.22. Part of the successful Blue Ribbon chain that appears multiple times in top revenue generators.

Shake Shack

Dominant market position with highest revenue ($3,579.53) and solid 4.17 rating. Essential partner for promotional campaigns.

Strategic Recommendations

Revenue Growth Strategies

- Promote bundle deals & free delivery to encourage orders above $20.
- Feature high-performing restaurants in targeted promotions.

Operational Improvements

- Implement dynamic delivery pricing for peak weekday hours.
- Optimize order queuing for cuisines with long preparation times.

Customer Retention & Loyalty

- Launch a loyalty program for high-spending customers.
- Enhance real-time order tracking & estimated delivery time transparency.

Action Plan Implementation

1

Phase 1: Revenue Optimization

- Implement bundle deals promotion for orders above $20
- Create featured section for high-performing restaurants like Shake Shack

2

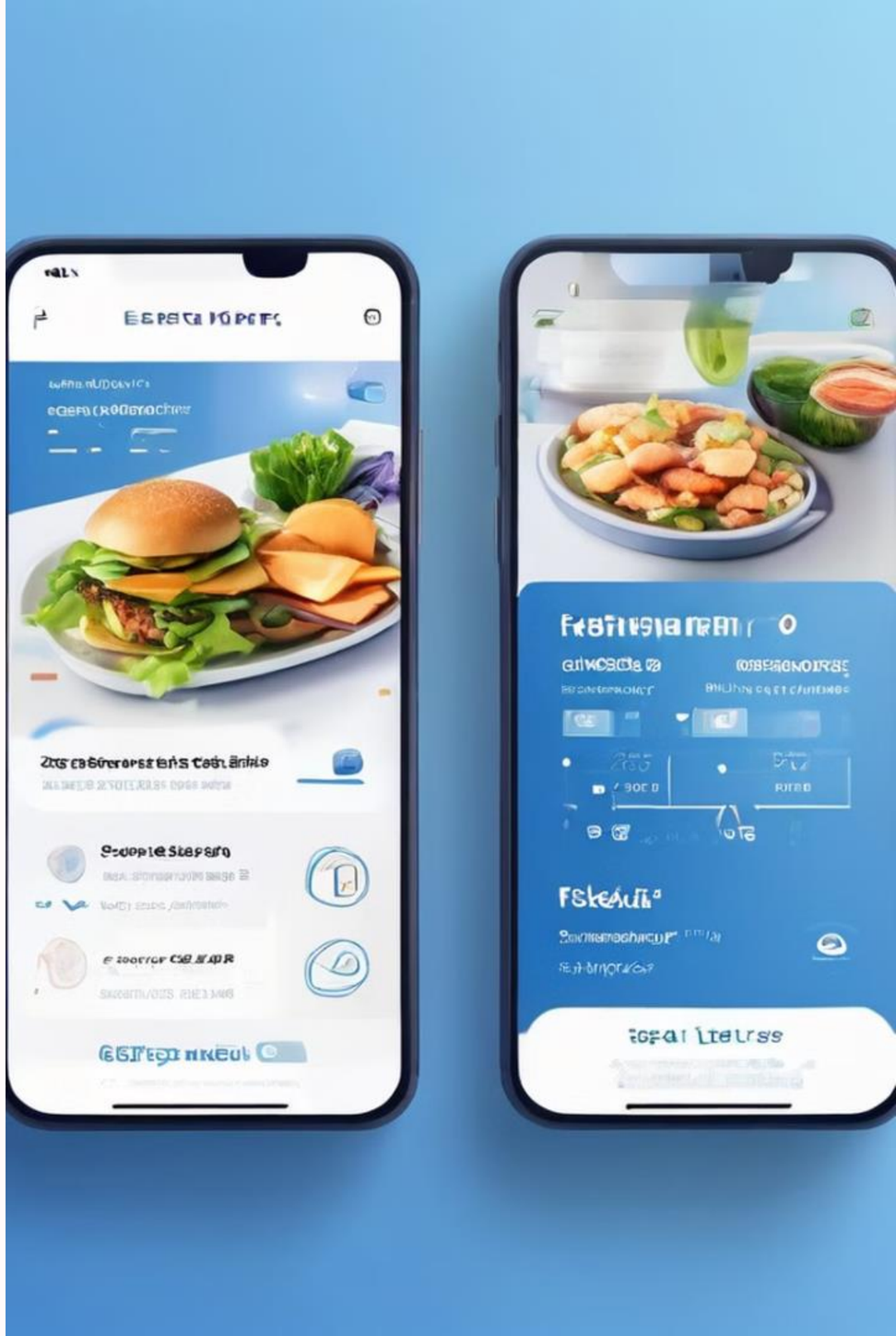
Phase 2: Delivery Efficiency

- Roll out dynamic pricing model for weekday peak hours
- Develop specialized queuing system for Japanese & American cuisine orders

3

Phase 3: Customer Experience

- Launch loyalty program targeting high-value customers
- Upgrade app with enhanced real-time tracking features



Business Problem Overview and Solution Approach

- Please define the problem
- Please mention the solution approach / methodology

***Note**: You can use more than one slide if needed*

The Business Problem

FoodHub, an online food delivery platform, aims to optimise its revenue streams, improve delivery efficiency, and enhance customer satisfaction.

The key challenges identified include:

Revenue Optimization:

- Which order price ranges contribute the most to revenue?
- How do restaurant performance and cuisine type impact earnings?

Delivery Efficiency:

- Do weekdays vs. weekends affect delivery time?
- Which cuisines take longer to prepare and deliver?

Customer Satisfaction & Retention:

- Does delivery speed impact customer ratings?
- Which top-rated restaurants should be promoted?

Methodology

A data-driven approach using Exploratory Data Analysis (EDA) and statistical insights was implemented.

Data Cleaning & Preprocessing

- Missing values in ratings were imputed using median ratings per restaurant.
- Order costs and revenue were categorized into low, mid, and high-value segments.

Descriptive & Statistical Analysis

- Revenue and order trends were analyzed by cuisine type, restaurant name, and spending habits.
- Delivery time distribution was evaluated across weekdays vs. weekends and different cuisines.
- Customer ratings were compared across different service conditions (fast vs. slow deliveries).

Data Visualization & Correlation Analysis

- Scatter plots, box plots, and heatmaps were used to identify relationships between key business variables.
- A correlation heatmap was generated to understand the impact of cost, ratings, and delivery time.

Data Dictionary

order_id	Unique ID of the order
customer_id	ID of the customer who ordered the food
restaurant_name	Name of the restaurant
cuisine_type	Cuisine ordered by the customer
cost_of_the_order	Cost of the order
day_of_the_week	Indicates whether the order is placed on a weekday or weekend
rating	Rating given by the customer out of 5
food_preparation_time	Time (in minutes) taken by the restaurant to prepare the food
delivery_time	Time (in minutes) taken by the delivery person to deliver the food package

Data Overview

- Please add the data overview
- Please add answers for all question from 1 till 5

***Note**: You can use more than one slide if needed*

Data Structure Analysis

1

Question 1 Dataset Size

The dataset has 1898 rows and 9 columns, providing a substantial amount of data for analysis.

2

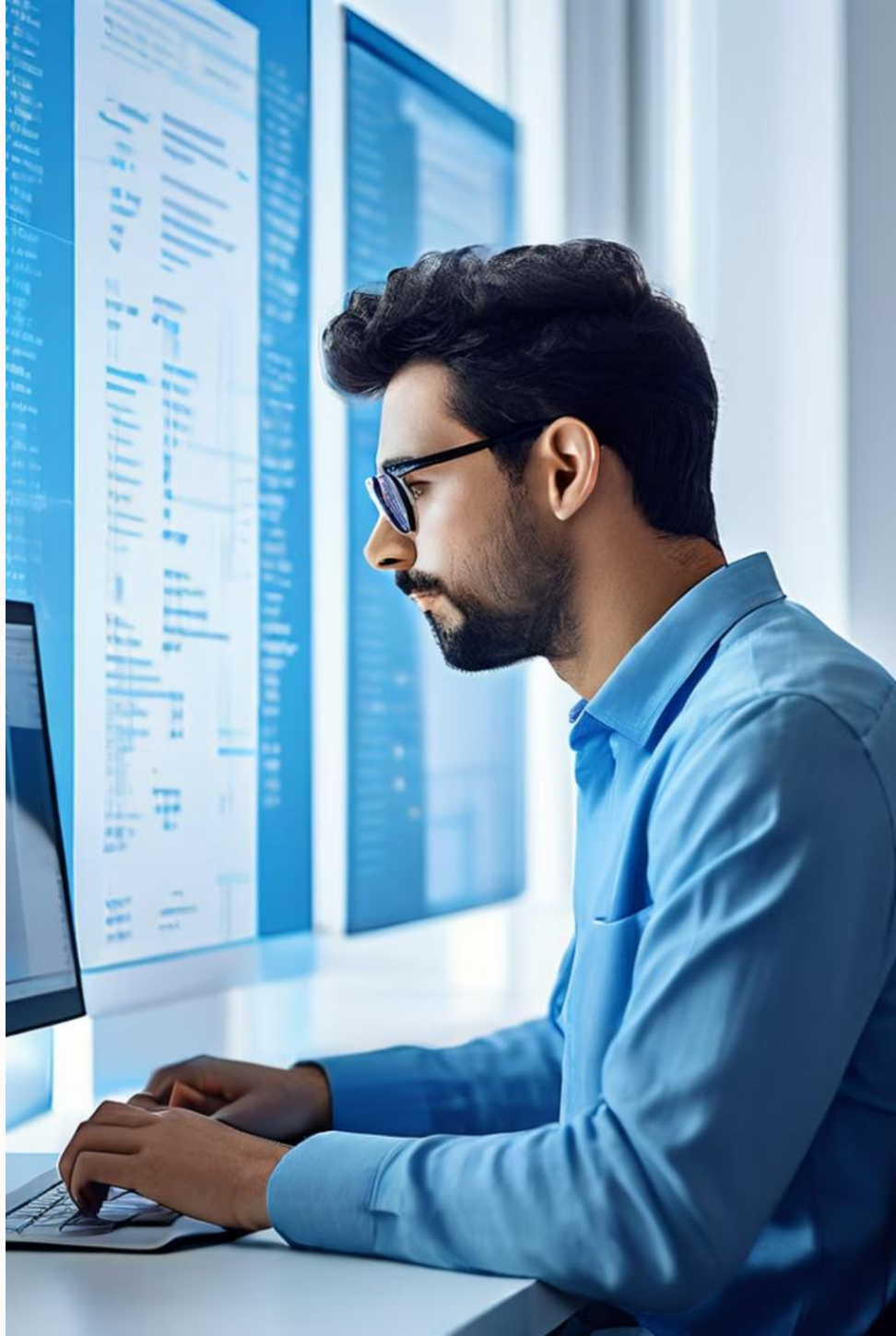
Question 2 Data Types

The dataset contains a mix of numerical data (int64, float64) and categorical data (object), requiring different analytical approaches.

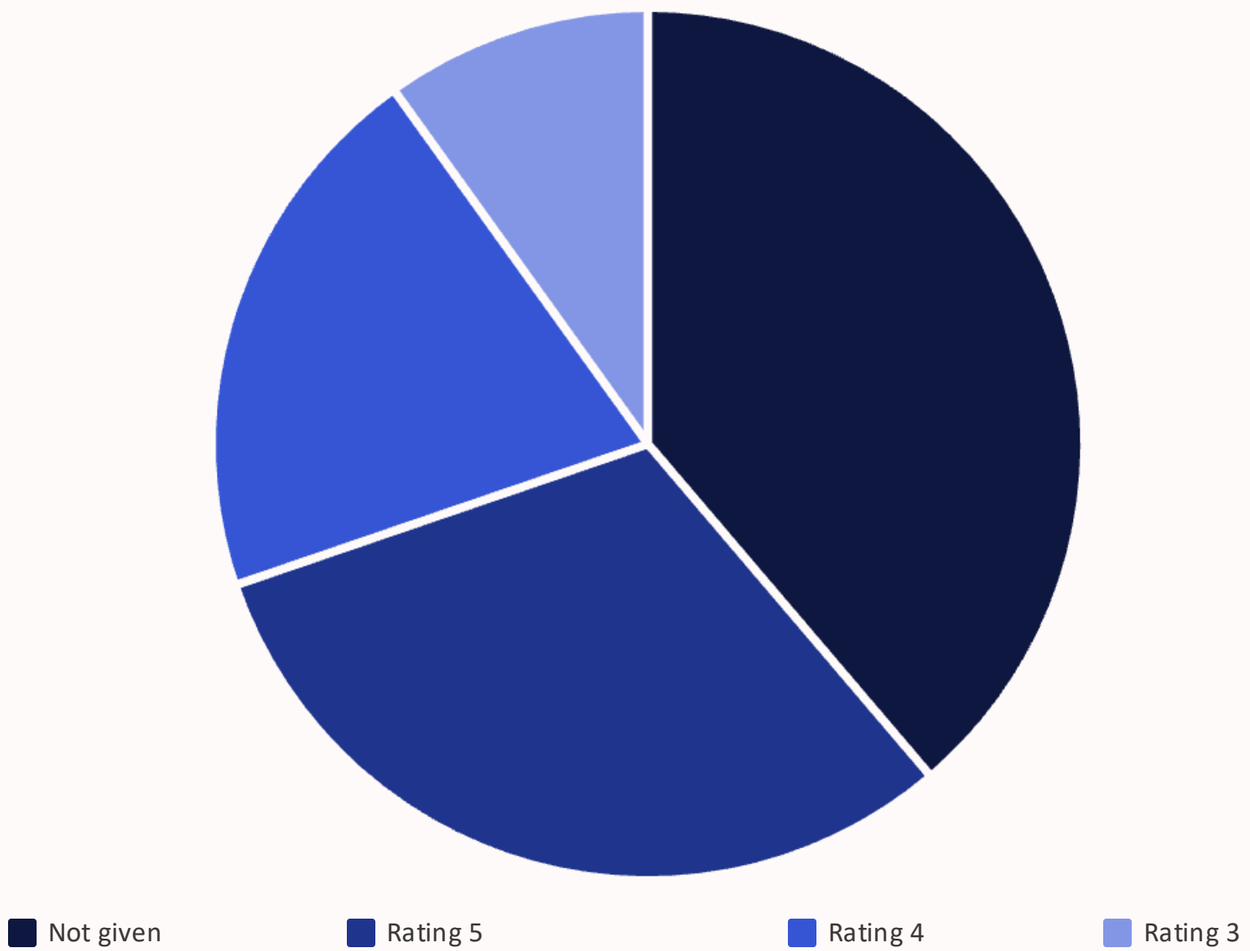
3

Question 3 Complete Records

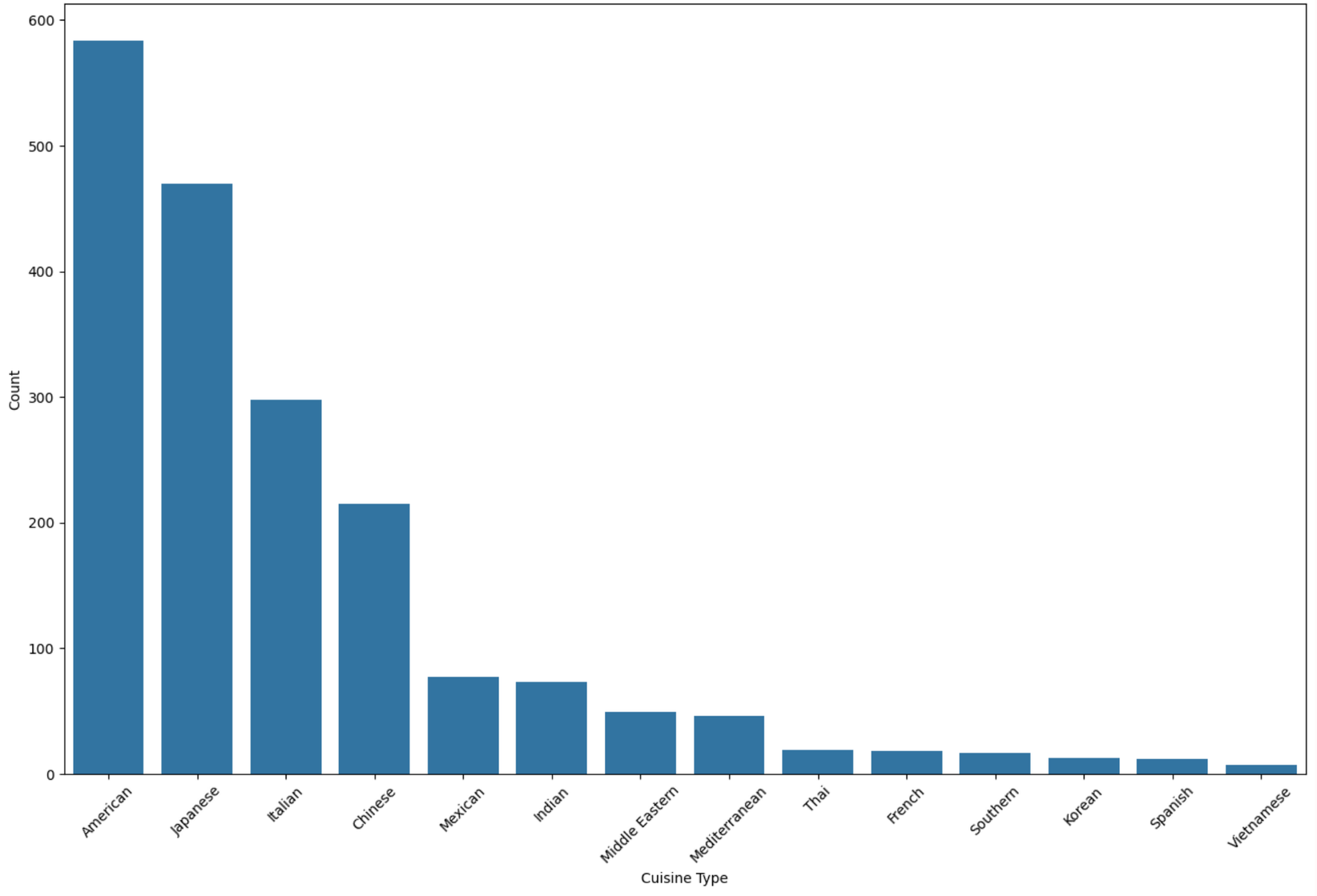
All columns have 1,898 non-null values, meaning the dataset has no missing values. This simplifies preprocessing but still requires verification for incorrect or inconsistent data entries.



Rating Distribution Analysis



A significant portion (38.78%) of the ratings are labeled as "Not given", meaning a large portion of customers did not provide a rating. Among those who did rate, 30.98% gave 5 stars, 20.34% gave 4 stars, and only 9.91% rated their experience as average (3 stars). This suggests that customers who do provide ratings tend to be satisfied with their experience.



The dataset is imbalanced with a certain preference for specific cuisines. American cuisine dominates with 30.77% of orders, followed by Japanese (24.76%), Italian (15.70%), and Chinese (11.33%). The remaining cuisines each represent less than 5% of the total orders. This distribution could be related to the demographics of the area or pricing preferences.

Question 4 Food Preparation Time Analysis

20

Minimum Time

The fastest food preparation time in minutes

27.37

Average Time

The mean food preparation time in minutes

35

Maximum Time

The longest food preparation time in minutes

The food preparation time ranges from 20 minutes to 35 minutes, with an average of 27.37 minutes. This indicates that restaurants generally take between 20-35 minutes to prepare orders once they are placed. Understanding these preparation times can help set customer expectations and optimize delivery scheduling.

Question 5: Missing Ratings Analysis

Total Missing Ratings

736 orders (38.78%) were initially not given ratings, indicating that many customers did not provide feedback. This significant portion of missing data required treatment for comprehensive analysis.

Top Restaurants with Missing Ratings

- Shake Shack: 86 missing ratings
- The Meatball Shop: 48 missing ratings
- Blue Ribbon Sushi: 46 missing ratings
- Blue Ribbon Fried Chicken: 32 missing ratings
- Parm: 29 missing ratings

Univariate Analysis

- Please mention regarding univariate analysis for all columns
- Please add answers for all question from 6 till 11

***Note**: You can use more than one slide if needed*

Customer Order Patterns

Unique Orders

Each order has a unique order ID (1,898 total), meaning the dataset does not contain duplicated orders.

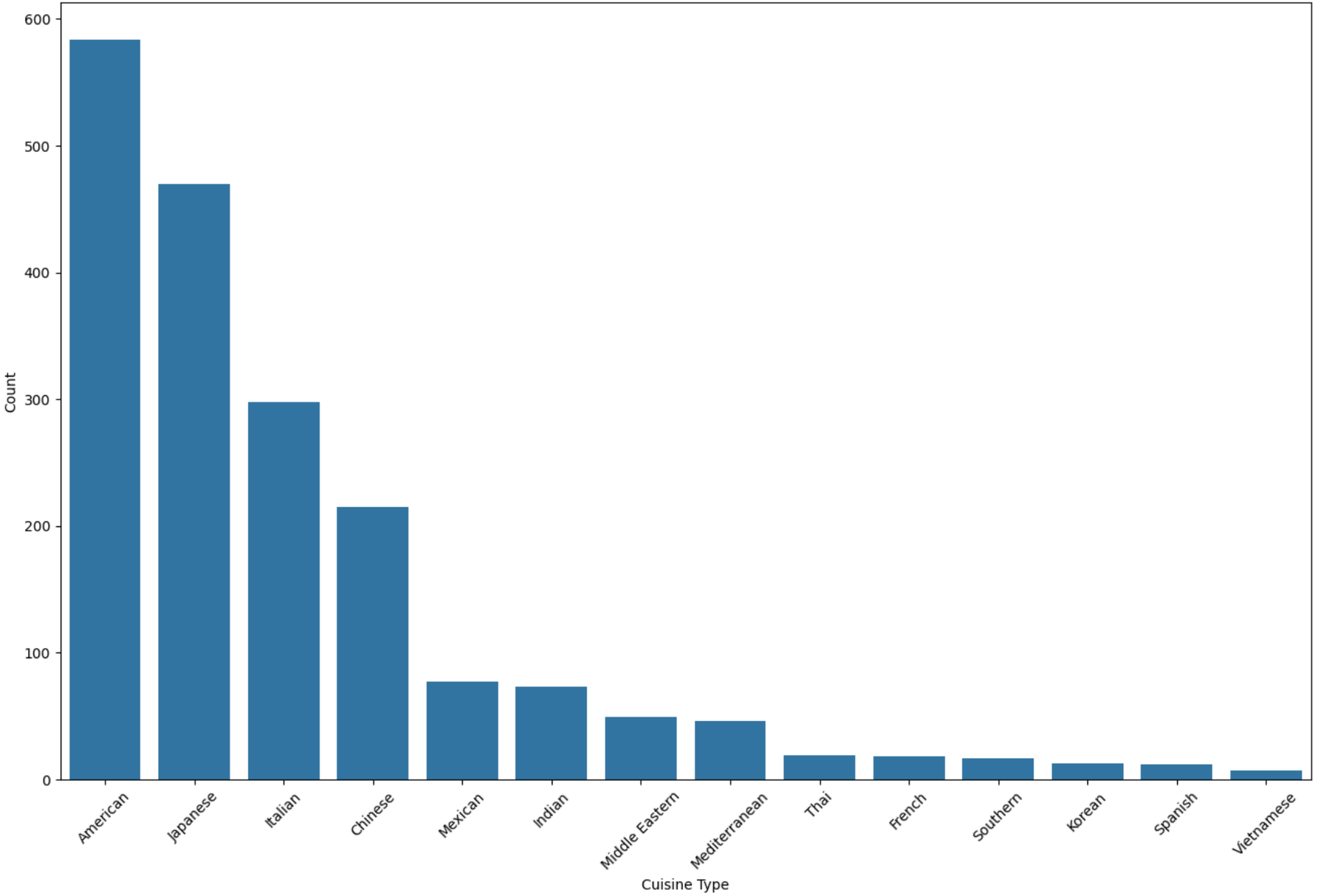
Customer Base

1,200 unique customers created 1,898 orders, suggesting approximately 1.58 orders per customer, indicating customer repeat business.

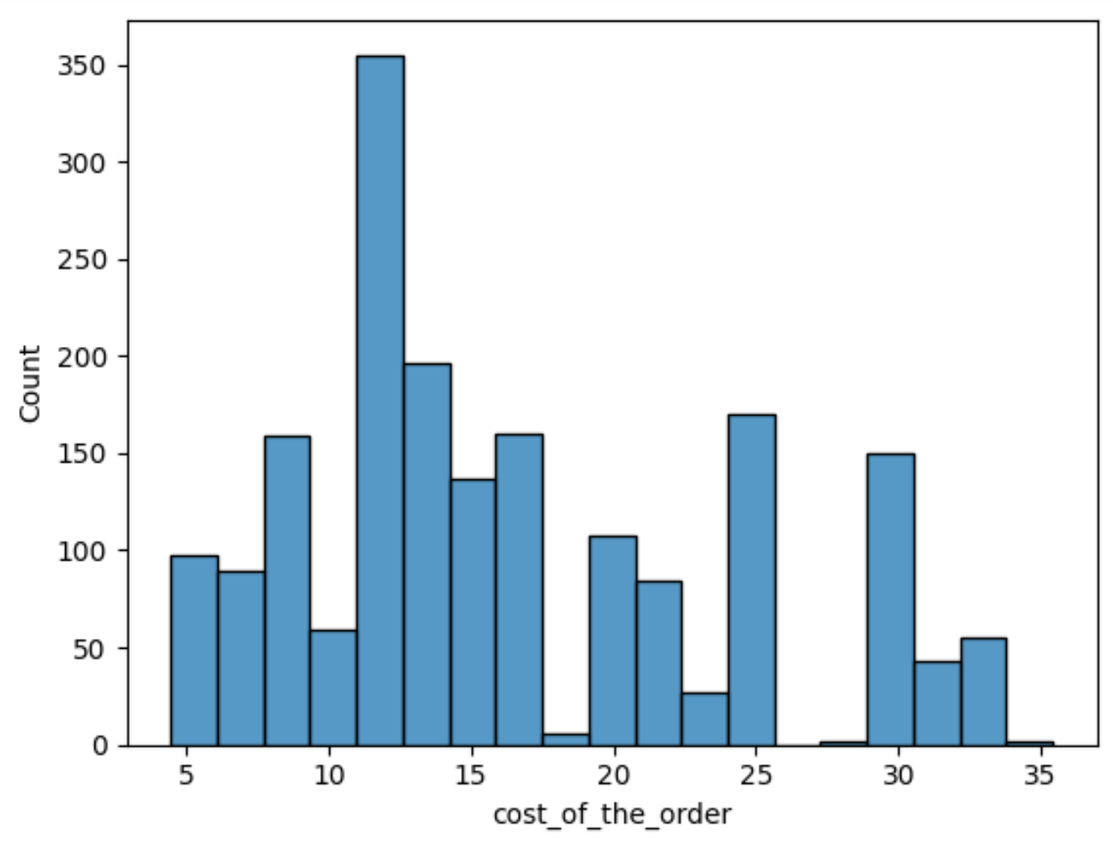
Restaurant Variety

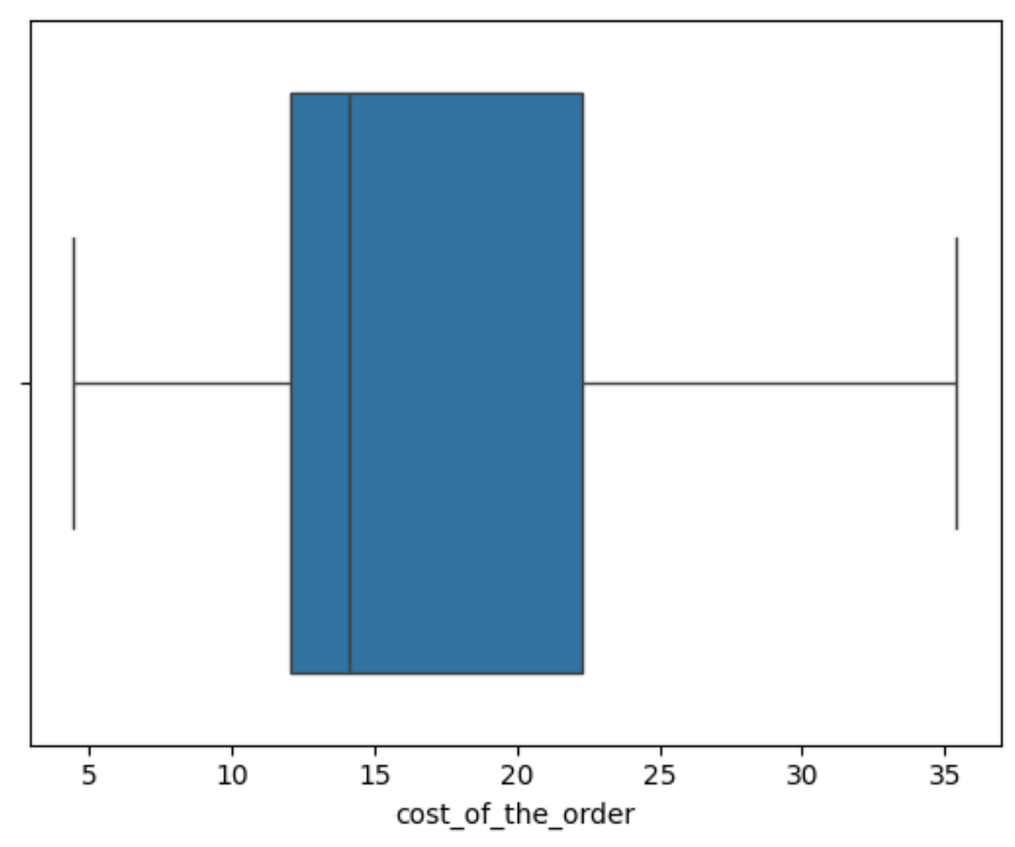
178 unique restaurants exist in the dataset, offering 14 different cuisine types to customers.

Cuisine types



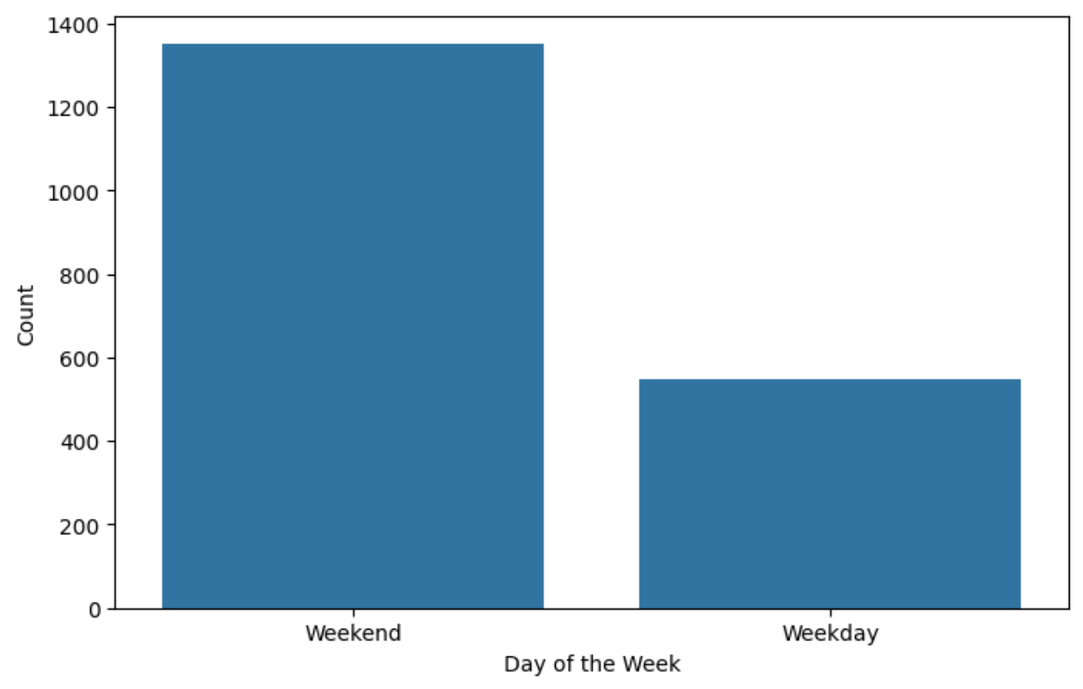
Order Cost Distribution





The order cost distribution is right-skewed (positively skewed), with the most common price point around $12. The distribution shows specific price points with higher frequencies, such as $12, $25, and $30. There are no significant outliers, with most costs falling within a reasonable range between $12 and $22.

Weekday vs. Weekend Orders



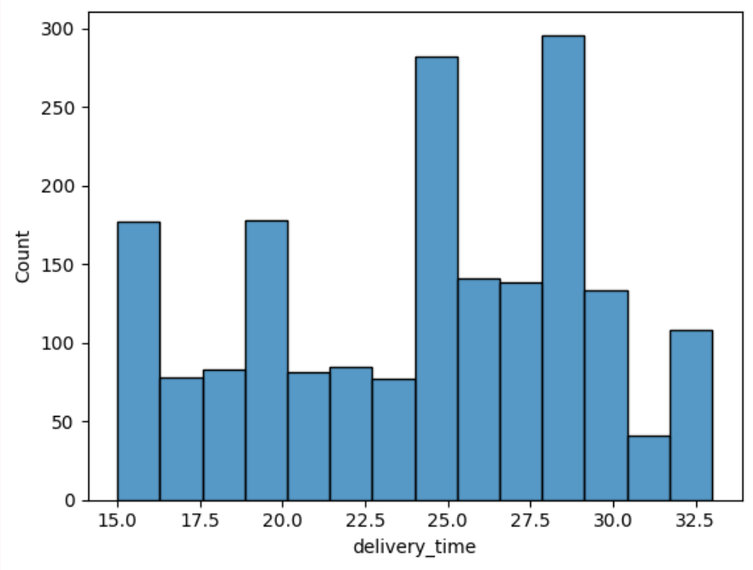
Weekend Orders

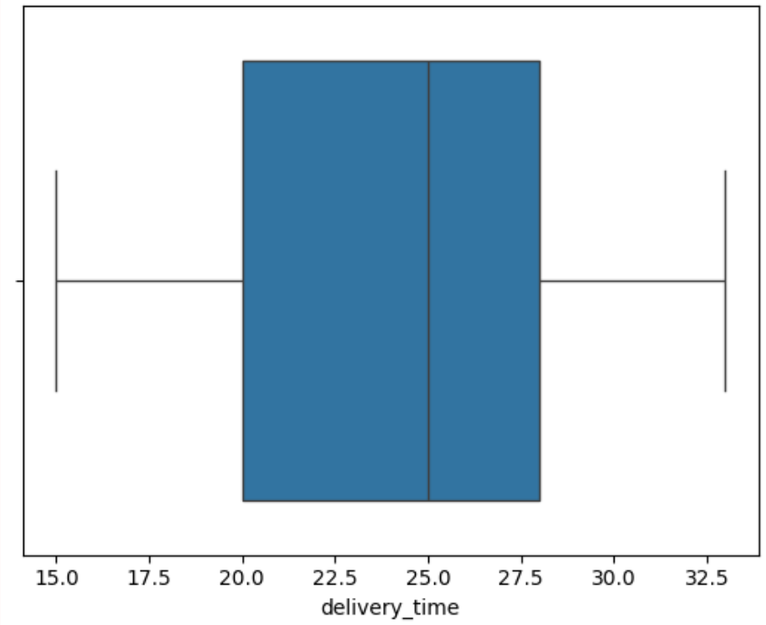
Most orders occur on weekends, with over twice as many orders as weekdays. This could be due to several reasons: people prefer takeaways for social activities, and some people prefer to cook for themselves and maintain a routine during weekdays.

Weekday Orders

Weekday orders are significantly fewer than weekend orders. This presents an opportunity for the business to implement promotional strategies to increase weekday order volume.

Delivery Time Analysis





1

Distribution Pattern

The most common delivery time is approximately 28-29 minutes. Other notable high counts include 15, 20, and 25 minutes. The data does not follow a normal distribution and is left-skewed (negatively skewed).

2

Consistency

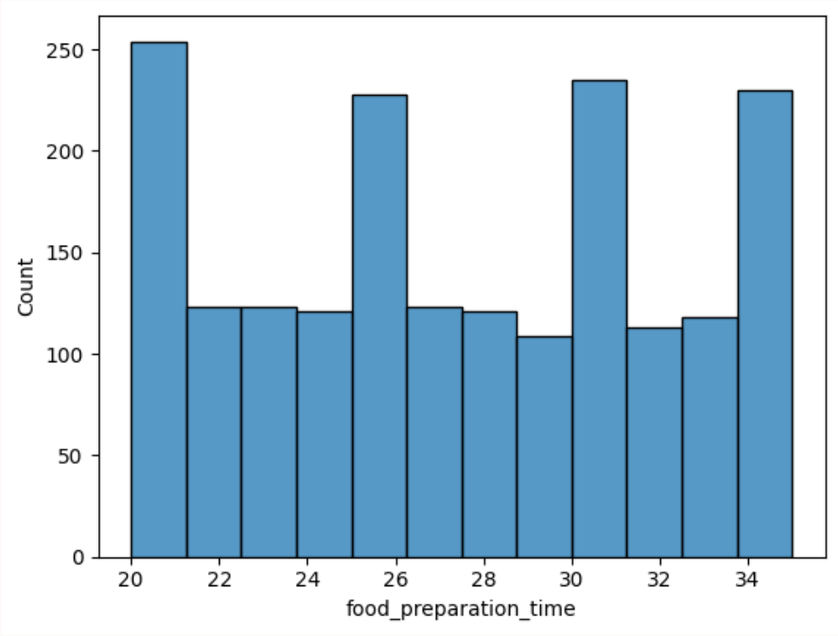
No outliers exist in the data, showing that delivery times remain consistent throughout. The distribution is slightly symmetrical with no strong skewness, though it does have some spikes at specific time intervals.

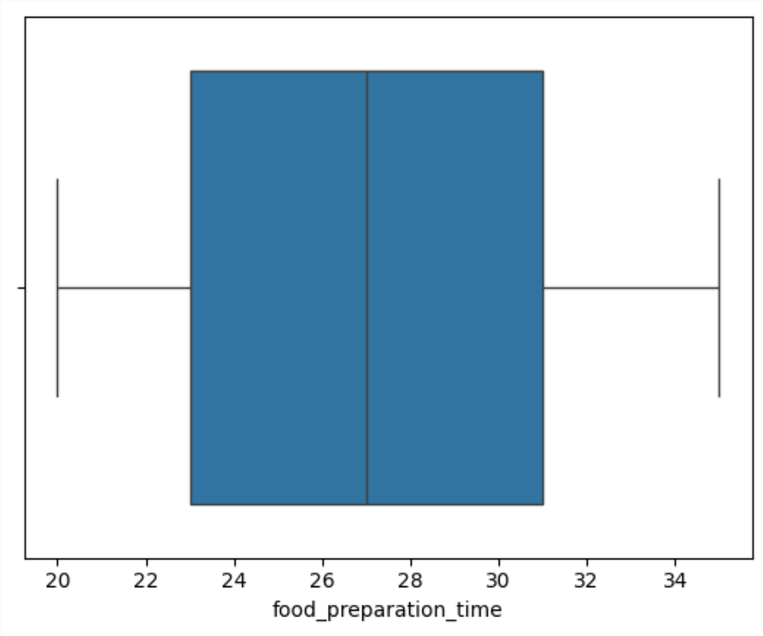
3

Delivery Timeframes

50% of deliveries take less than or equal to 25 minutes. 25% of deliveries are completed between 20-25 minutes, and another 25% are delivered between 25 and 28 minutes.

Food Preparation Time Patterns





Distribution Characteristics

The distribution appears uniform, with multiple peaks suggesting consistent preparation times across different order types. There is no strong skewness—the values are spread out between 20 to 35 minutes.

Common Preparation Times

The most frequent preparation times are around 21, 27, 30, and 34 minutes, which may indicate batch cooking or standard prep times for specific meals.

Median and Range

The median preparation time is around 27 minutes, meaning half of all orders take this time or less. 50% of the IQR falls between 23 and 31 minutes, with no major outliers.

Total Order Time Analysis

1. **Order Placement**

   Customer places order through the FoodHub app

2. **Food Preparation**

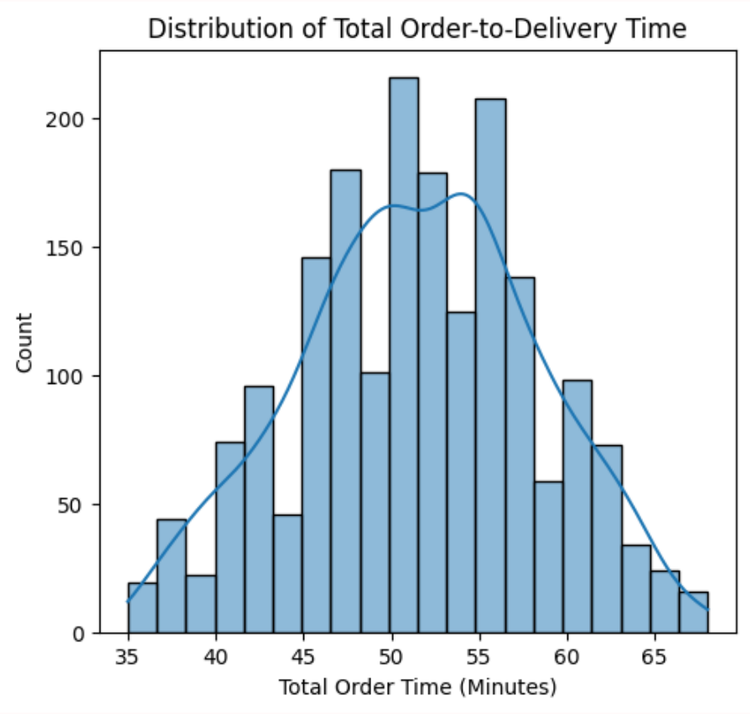
   Restaurant prepares the food (average 27.37 minutes)

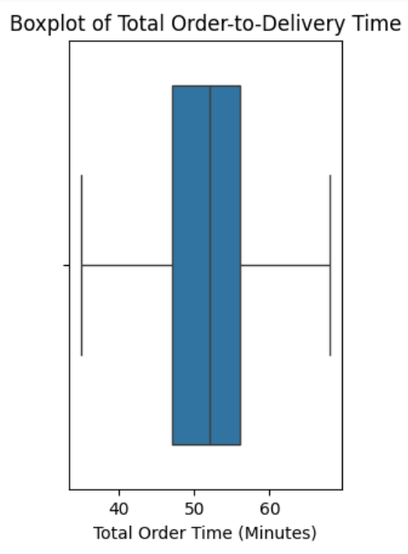
3. **Delivery Process**

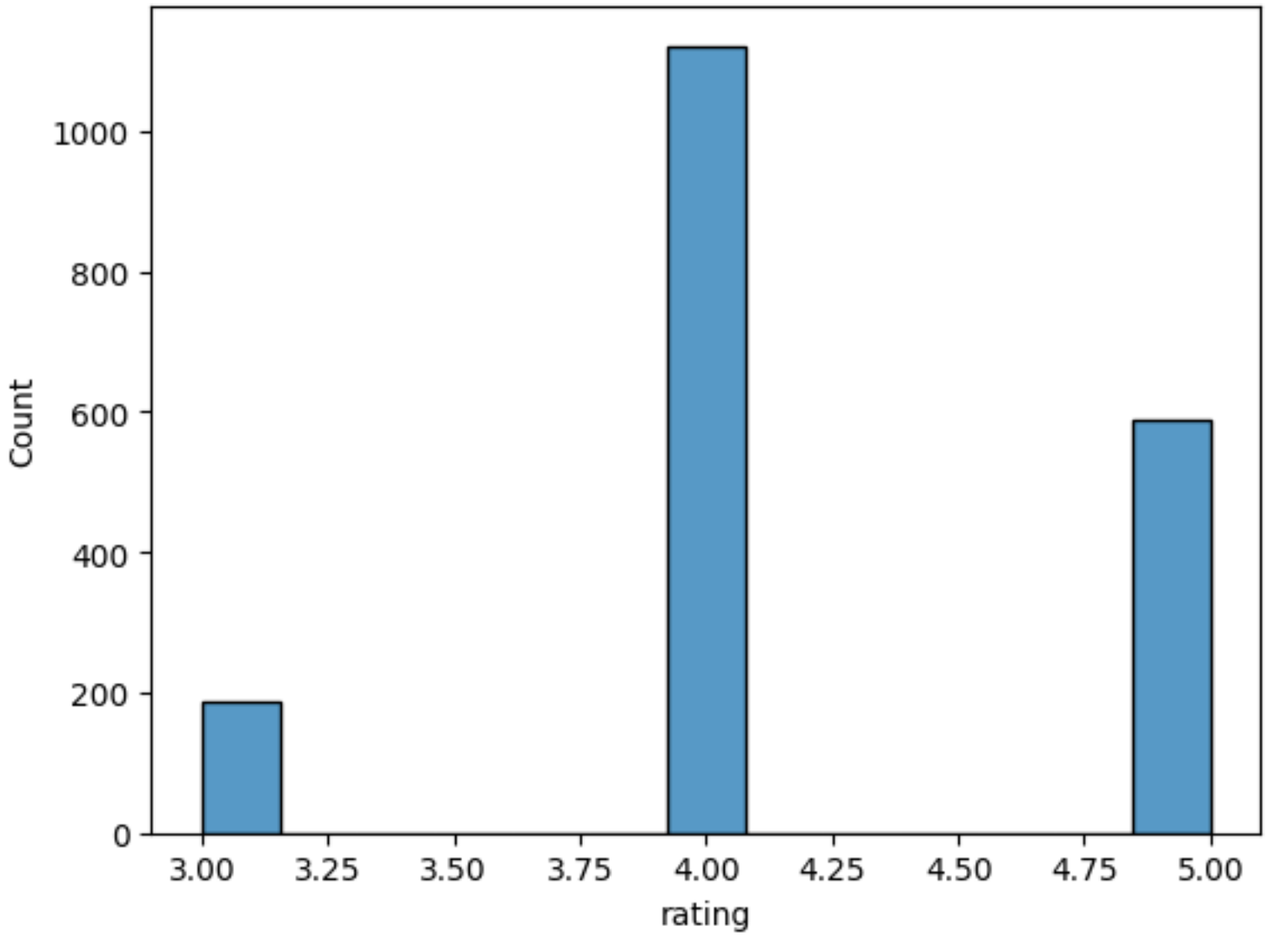
   Delivery person transports food to customer (average 24.16 minutes)

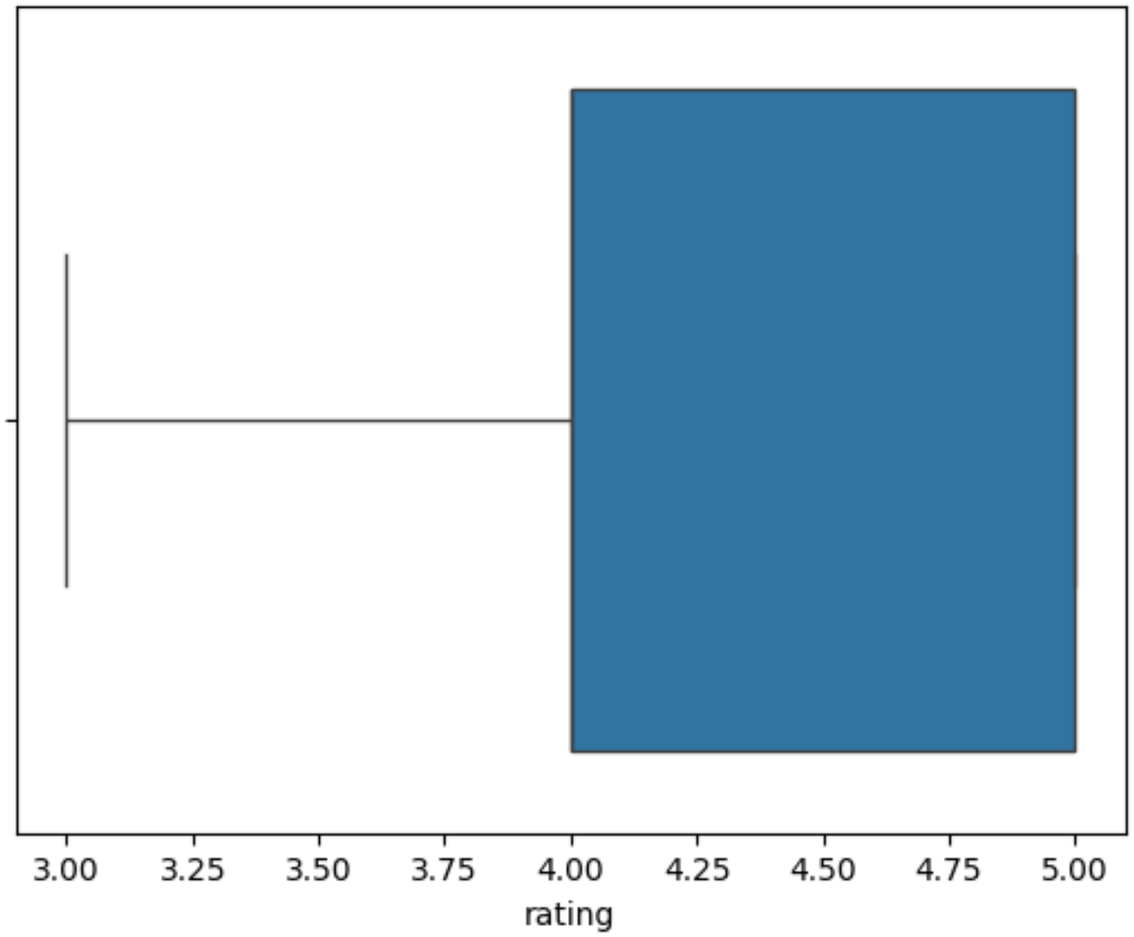
4. **Total Experience**

   Complete process from order to delivery (average 51.53 minutes)





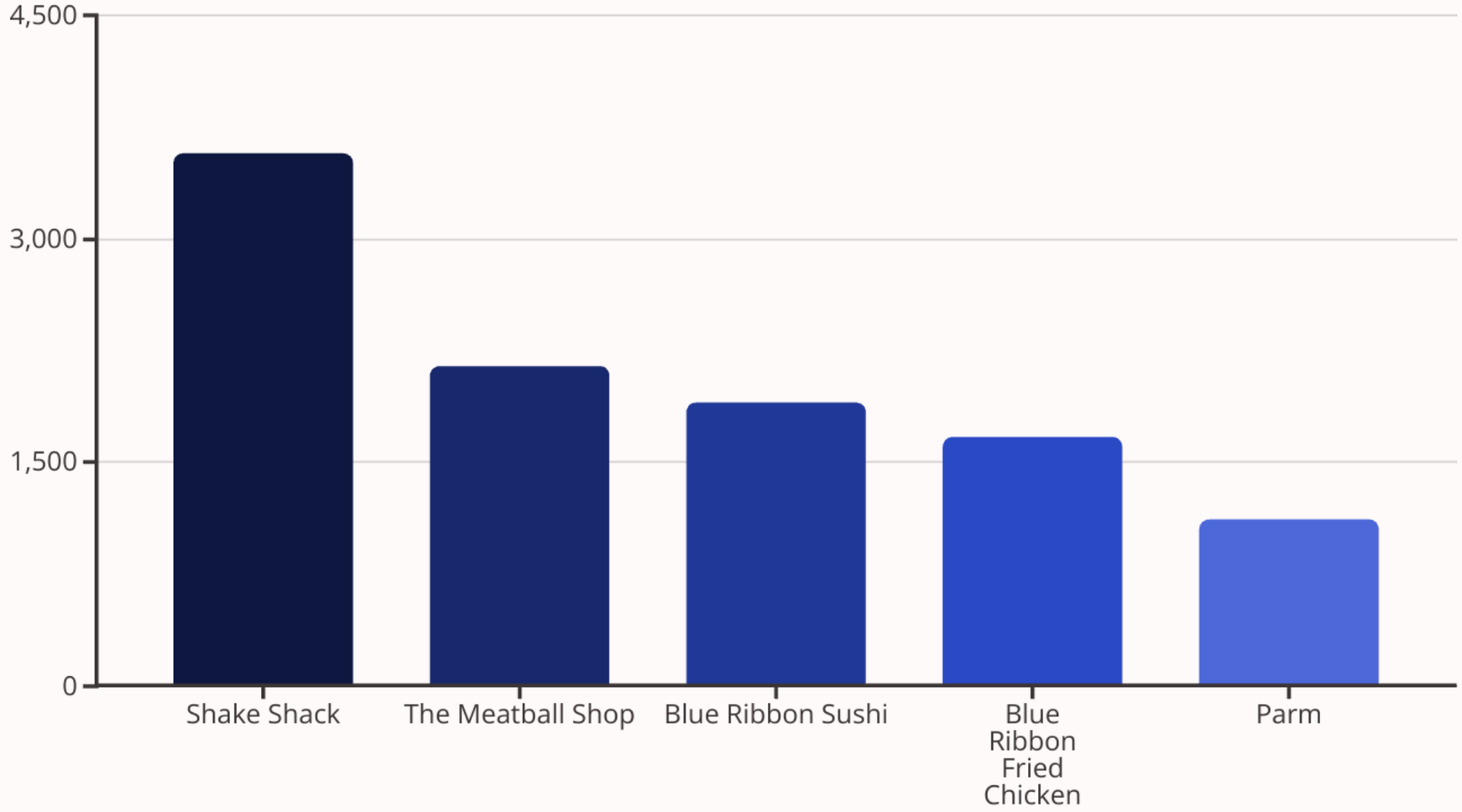




Observations on ratings.

- The distribution of ratings is heavily skewed towards positive feedback, which may indicate customer satisfaction or possible bias in ratings (e.g., fewer negative reviews).
- The limited variation in ratings could suggest that customers either give high ratings or do not rate at all.
- Since missing ratings were imputed, it is essential to ensure that my treatment did not inflate the average ratings too much, potentially skewing the overall perception of customer satisfaction.
- Im confident that we did not do that.

Question 7 Top Revenue-Generating Restaurants



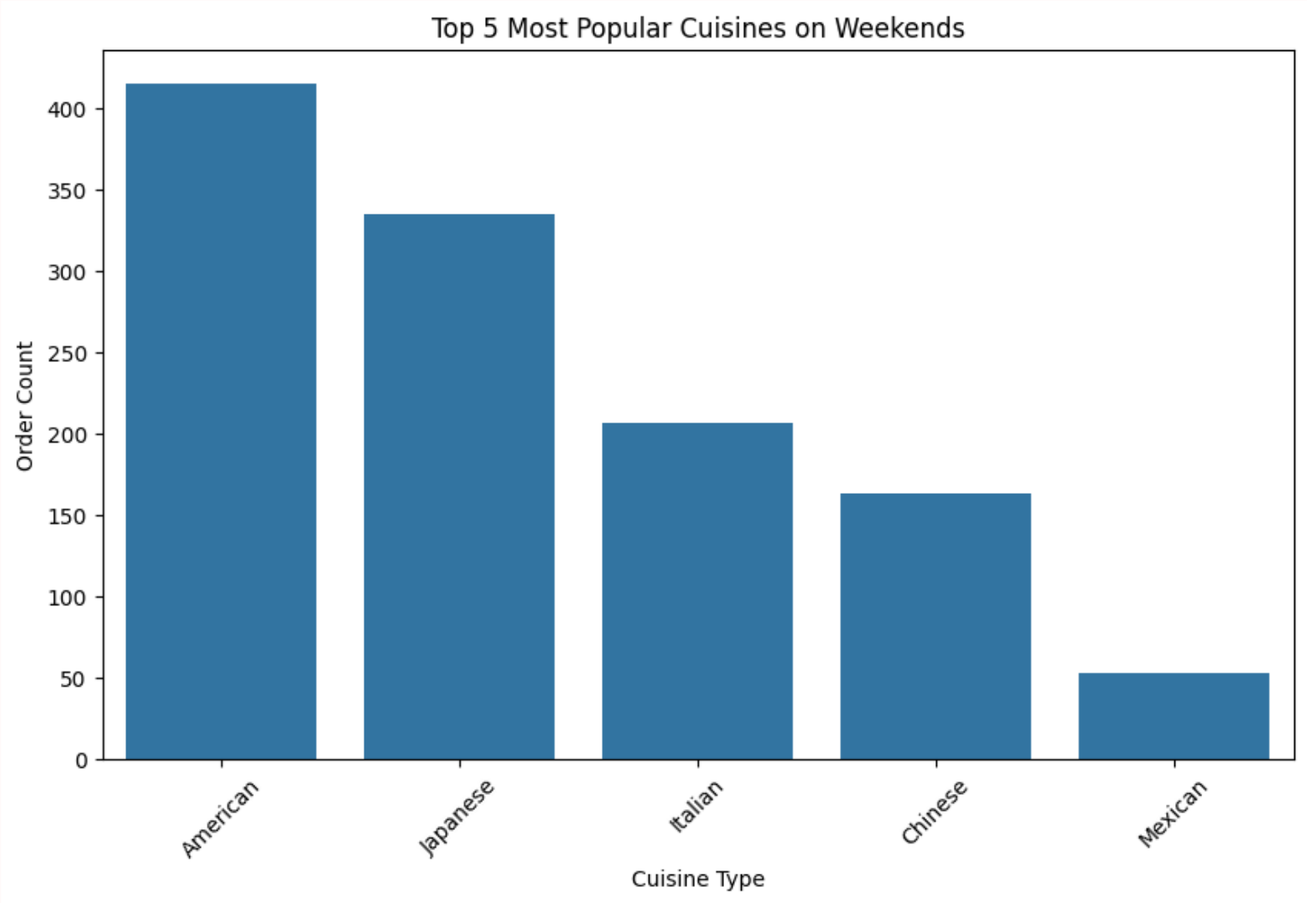
Shake Shack significantly outperforms all other restaurants, generating $3,579.53, nearly 67% more than the second-highest revenue generator. The Blue Ribbon chain (Sushi, Fried Chicken, and Sushi Bar & Grill) appears multiple times in the top 10, suggesting brand loyalty, strong market presence, or menu pricing contributing to higher revenue.

Question 8: Which is the most popular cuisine on weekends?

American 415

Observations.

- America stills ranks highest even on weekends; however, the total number of orders on weekends is close to the total orders of the week, which was approximately 600.
- This creates promotional opportunities for our business. We can provide customers on weekdays to increase their order volume.



Question 9 What percentage of the orders cost more than 20 dollars?

Observations:
29% of the total amount of orders cost more than $20.

Question 10: What is the mean order delivery time?

Weekday vs. Weekend Delivery Times

Weekday Delivery

Mean delivery time: 28.34 minutes

Weekday deliveries take longer, possibly due to traffic congestion, operational constraints, schools, businesses, and other factors that may contribute to additional time required for weekday deliveries.

Weekend Delivery

Mean delivery time: 22.47 minutes

Weekend deliveries are approximately 6 minutes faster on average than weekday deliveries. However, they show greater variability, possibly influenced by fluctuating demand patterns or staffing differences.

Question 11: Top three customers.

Most Frequent Customers

- Customer ID 52832: 13 orders, $225.80 total spent ($17.37 per order)
- Customer ID 47440: 10 orders, $158.18 total spent ($15.82 per order)
- Customer ID 83287: 9 orders, $139.31 total spent ($15.48 per order)

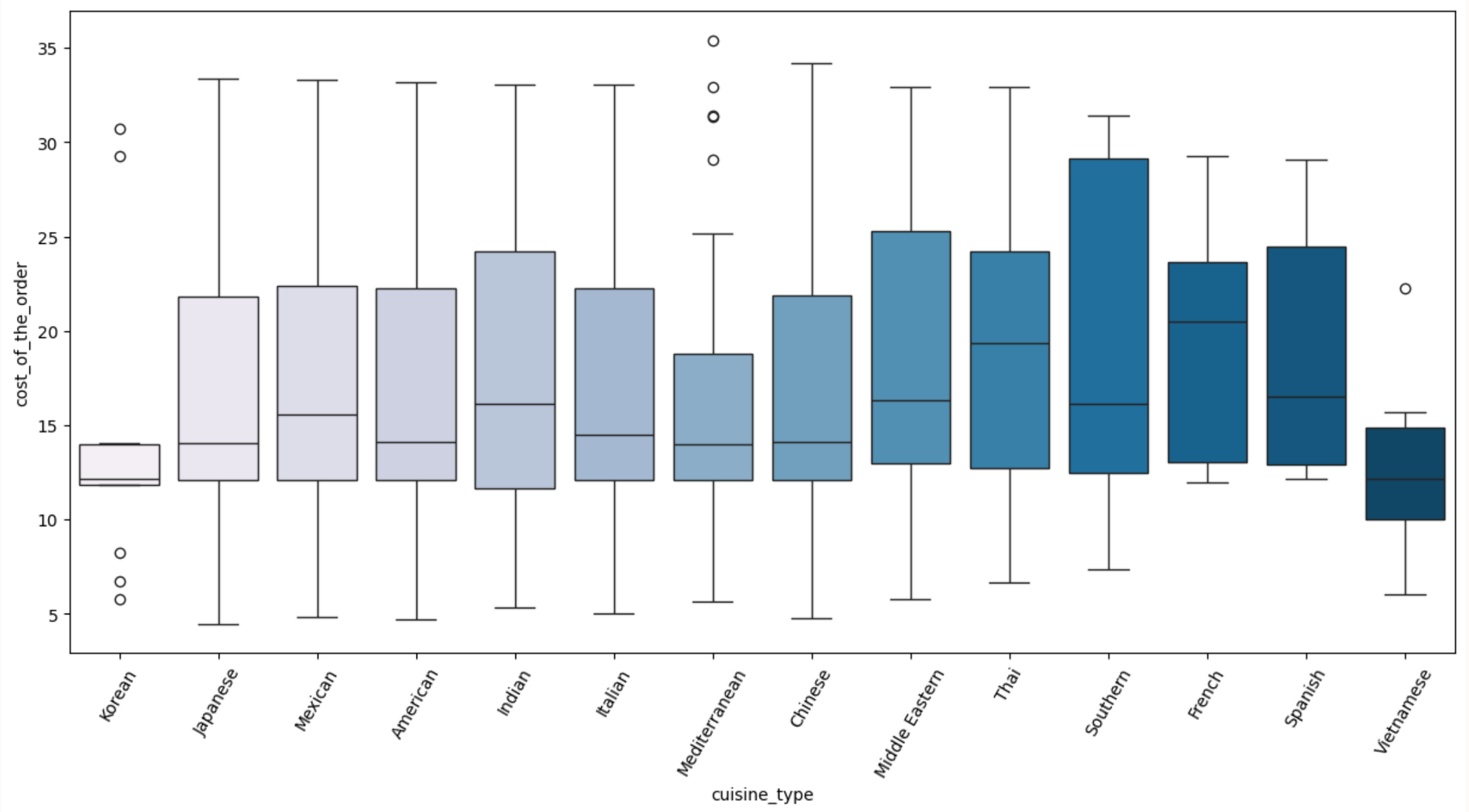
Business Implications

The top customers ordered a combined total of 32 orders, a small percentage of overall orders. This suggests an opportunity to implement a loyalty program to improve overall frequency and increase customer retention.

Multivariate Analysis

- Please mention regarding multivariate analysis between various columns
- Please add answers for all question from 12 till 16

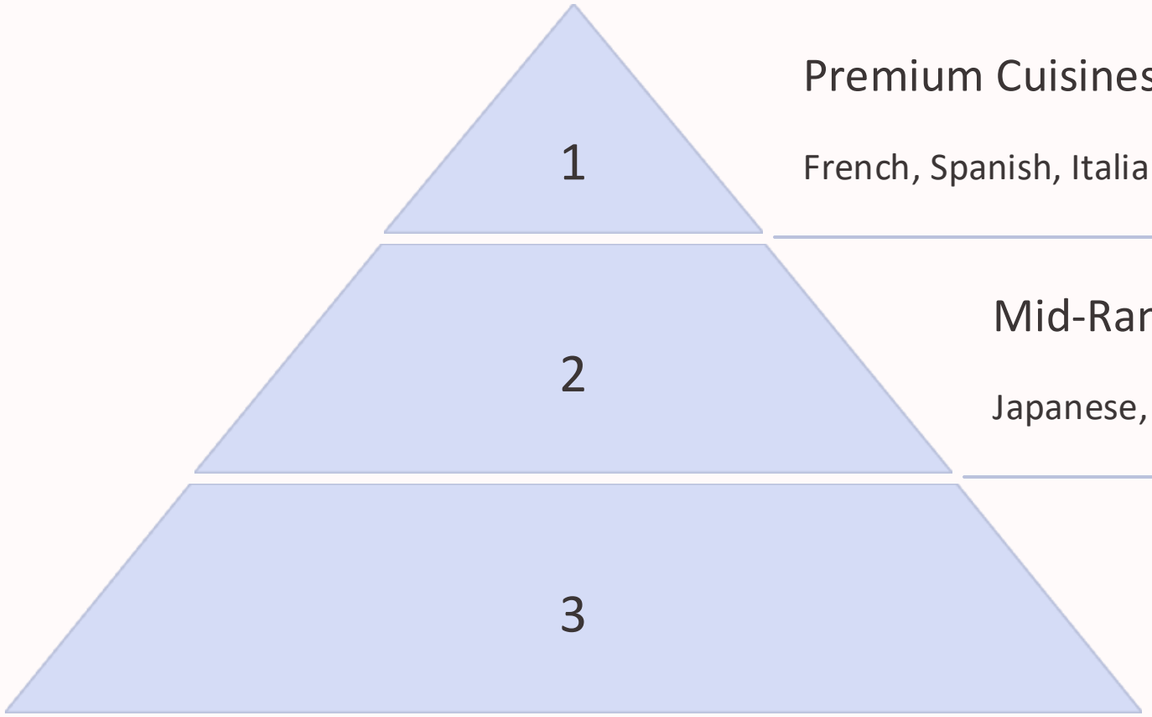
***Note**: You can use more than one slide if needed*



The median order costs vary significantly across cuisines. French, Spanish, and Italian cuisines have the highest median costs, while Vietnamese and Korean cuisines are more affordable with the lowest median prices.

Some cuisines show high variability in prices, particularly Southern, Spanish, Middle Eastern, Indian, and Thai, indicating they offer both budget and premium options.

Cuisine Cost Analysis



Premium Cuisines

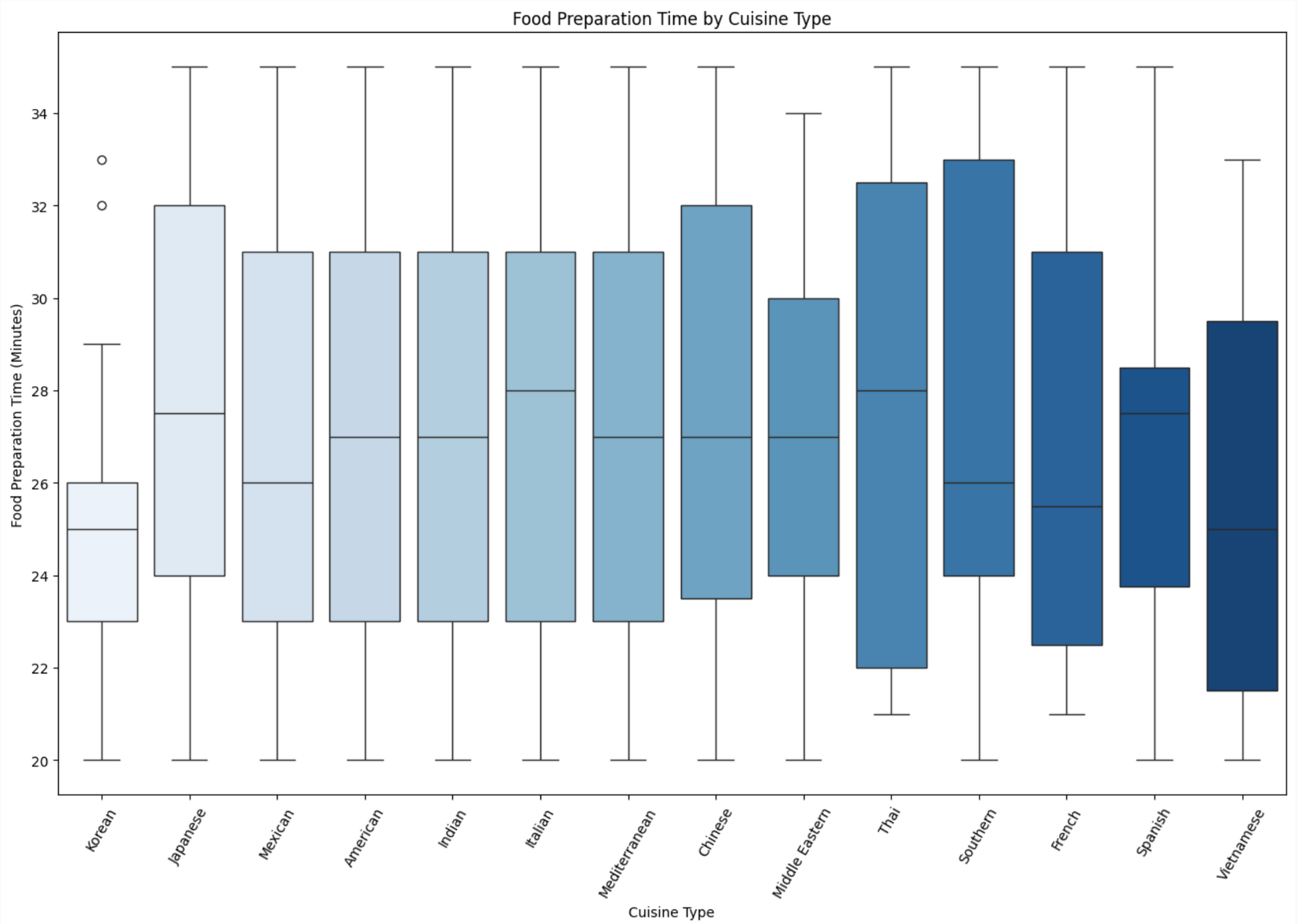
French, Spanish, Italian ($20+)

Mid-Range Cuisines

Japanese, American, Thai, Indian ($15-20)

Budget-Friendly Cuisines

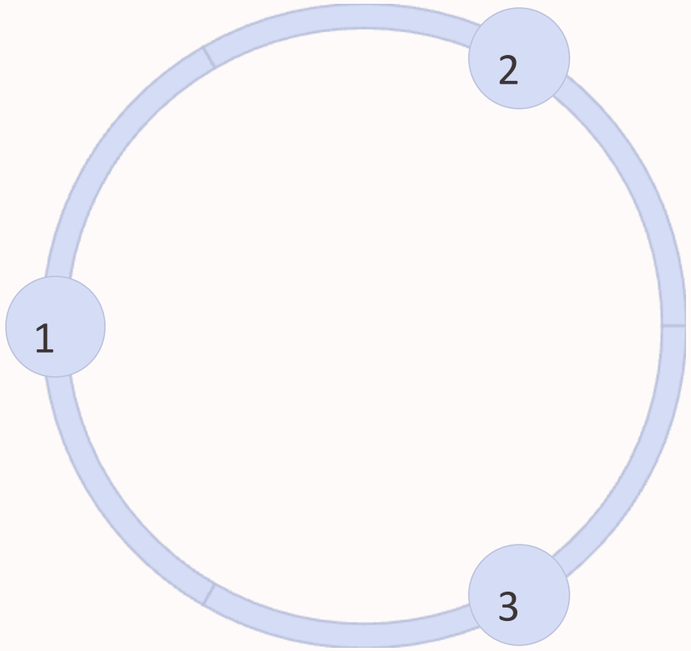
Vietnamese, Korean, Mediterranean ($10-15)


Food Preparation Time by Cuisine Type
Food Preparation Time (Minutes)
34
32
30
28
26
24
22
20
Korean
Japanese
Mexican
American
Indian
Italian
Mediterranean
Chinese
Middle Eastern
Thai
Southern
French
Spanish
Vietnamese
Cuisine Type

Cuisine Preparation Time Comparison

Quick Preparation (20-25 min)

Vietnamese and Korean cuisines have the shortest median preparation times, around 21-25 minutes. Mexican and Italian cuisines also have relatively lower preparation times.

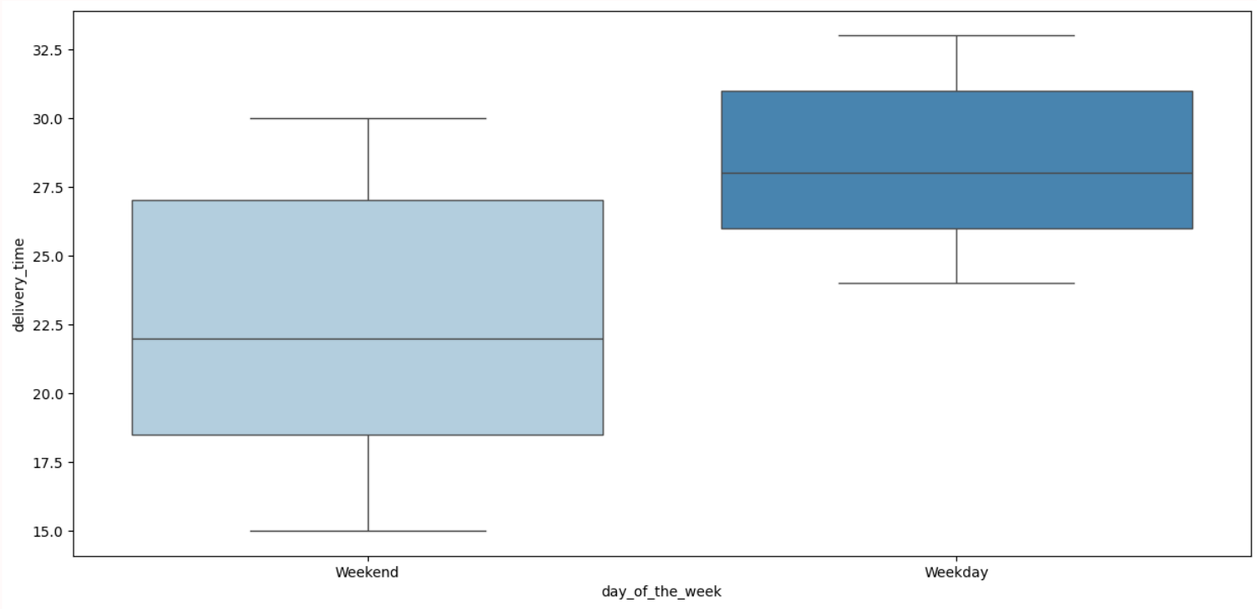


Medium Preparation (25-28 min)

American, Chinese, and Mediterranean cuisines fall in the middle range for preparation times.

Longer Preparation (28-30+ min)

French and Southern cuisines have the highest median preparation times, nearing 30 minutes. Japanese, Middle Eastern, and Thai cuisines also have longer preparation times.



Observations from the Box Plot:
Day of the Week vs. Delivery Time:

Weekday Deliveries Take Longer:

- The median delivery time on weekdays is higher than on weekends.
- The interquartile range (IQR) for weekdays is also narrower, indicating more consistent delivery times.

Weekend Deliveries Are Faster but More Variable:

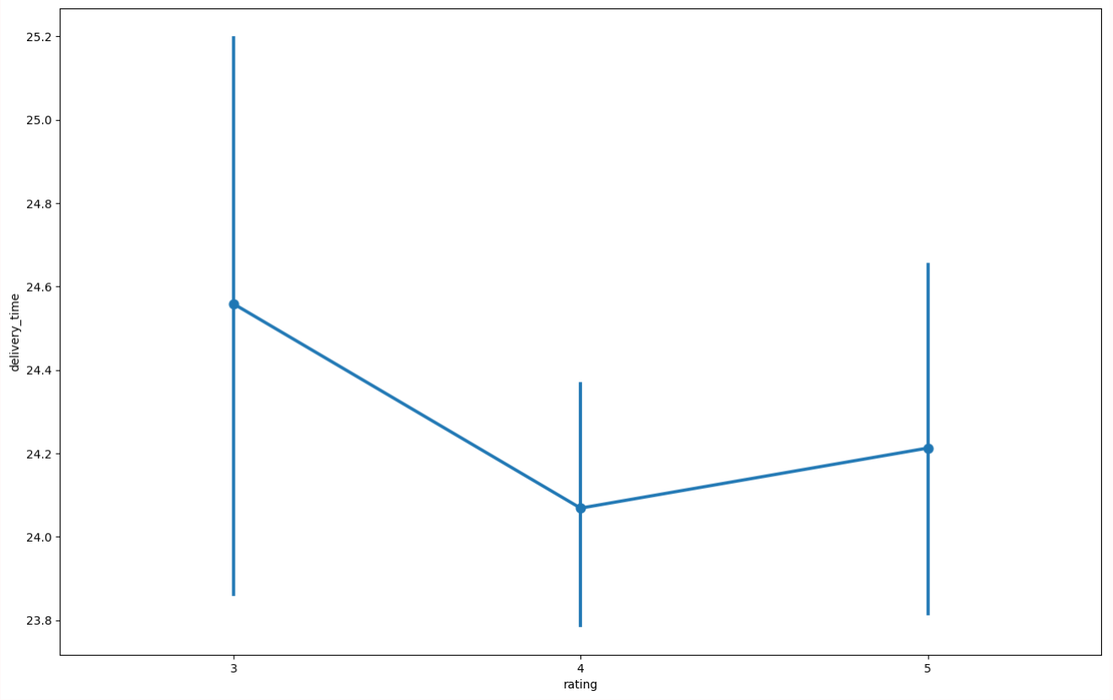
- The median delivery time is lower on weekends.
- The IQR is wider, suggesting greater variability in delivery times, possibly due to unpredictable demand surges.

Possible Operational Insights:

- Weekday deliveries may take longer due to higher traffic congestion or operational constraints; schools, businesses, and other factors may contribute to the additional time required for weekday deliveries.
- Weekend deliveries may have more variability, possibly influenced by fluctuating demand patterns or staffing differences.

Top 14 Revenue-Generating Restaurants: restaurant_name

1. Shake Shack 3579.53
2. The Meatball Shop 2145.21
3. Blue Ribbon Sushi 1903.95
4. Blue Ribbon Fried Chicken 1662.29
5. Parm 1112.76
6. RedFarm Broadway 965.13
7. RedFarm Hudson 921.21
8. TAO 834.50
9. Han Dynasty 755.29
10. Blue Ribbon Sushi Bar & Grill 666.62
11. Rubirosa 660.45
12. Sushi of Gari 46 640.87
13. Nobu Next Door 623.67
14. Five Guys Burgers and Fries 506.47



Rating vs. Delivery Time Analysis

No Strong Correlation

Delivery time alone is not the main driver of ratings. The average delivery time for orders with different ratings is extremely close, suggesting other factors influence customer satisfaction more significantly.

Balanced Expectations

Ratings are slightly higher for deliveries that are not the absolute fastest, possibly due to factors like food quality and service being more important than speed alone.

Quality Over Speed

While slow delivery could be a factor in lower ratings, it's not the main determinant of customer satisfaction. This suggests focusing on food quality and accuracy may be more important than minimizing delivery time.

Rating vs. Food Preparation Time

1

Minimal Impact

Similar to delivery time, food preparation time does not strongly influence customer ratings. This reinforces that time factors alone don't determine customer satisfaction.

2

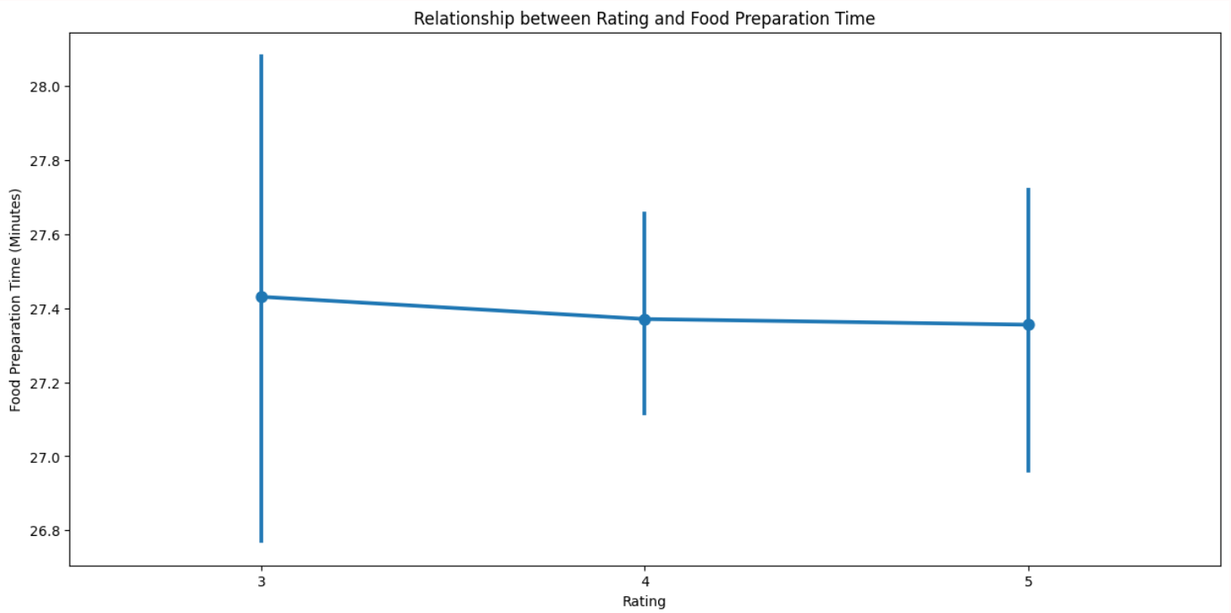
Highest Ratings

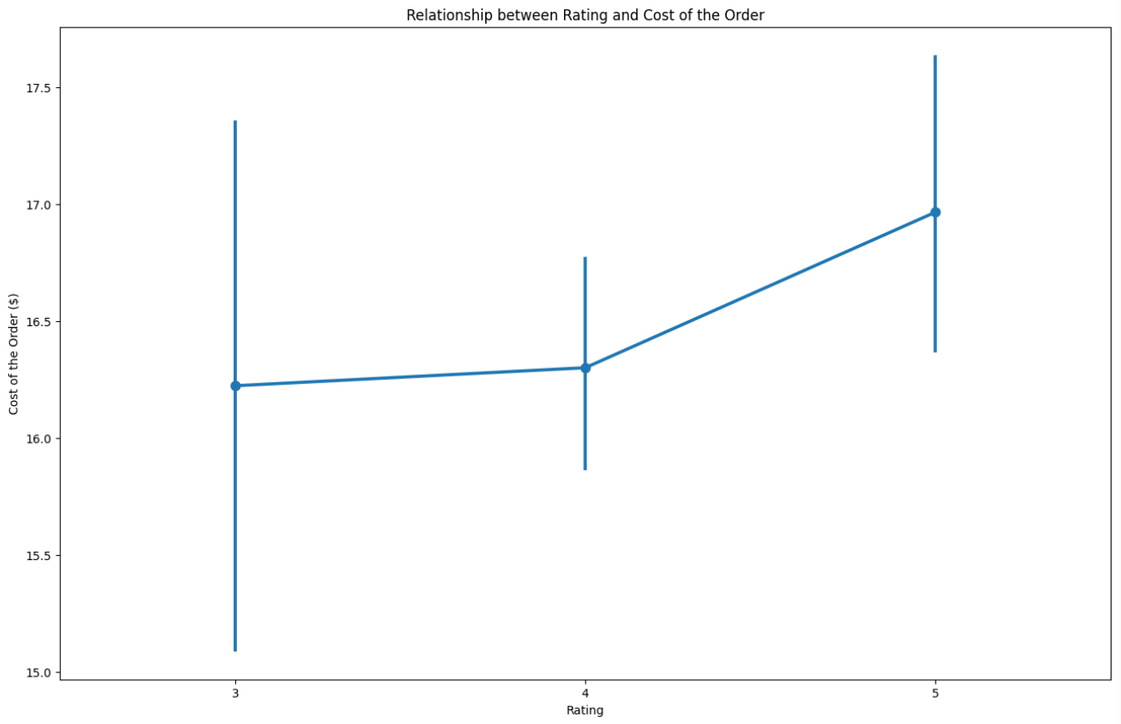
Orders with a rating of 4 have the shortest food preparation time on average, followed closely by those with a rating of 5.

3

Customer Expectations

The data suggests that customers may have reasonable expectations regarding preparation time, focusing more on other aspects of their experience when rating orders.

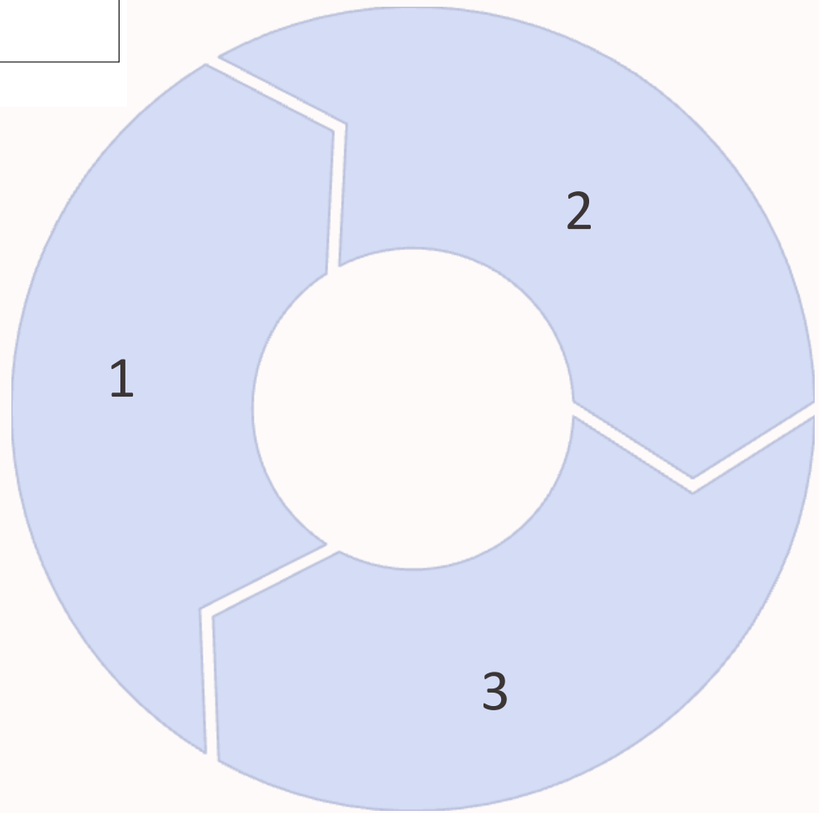




Rating vs. Order Cost Analysis

Higher Cost, Better Ratings

There is a slight upward trend in the cost of the order as ratings increase.



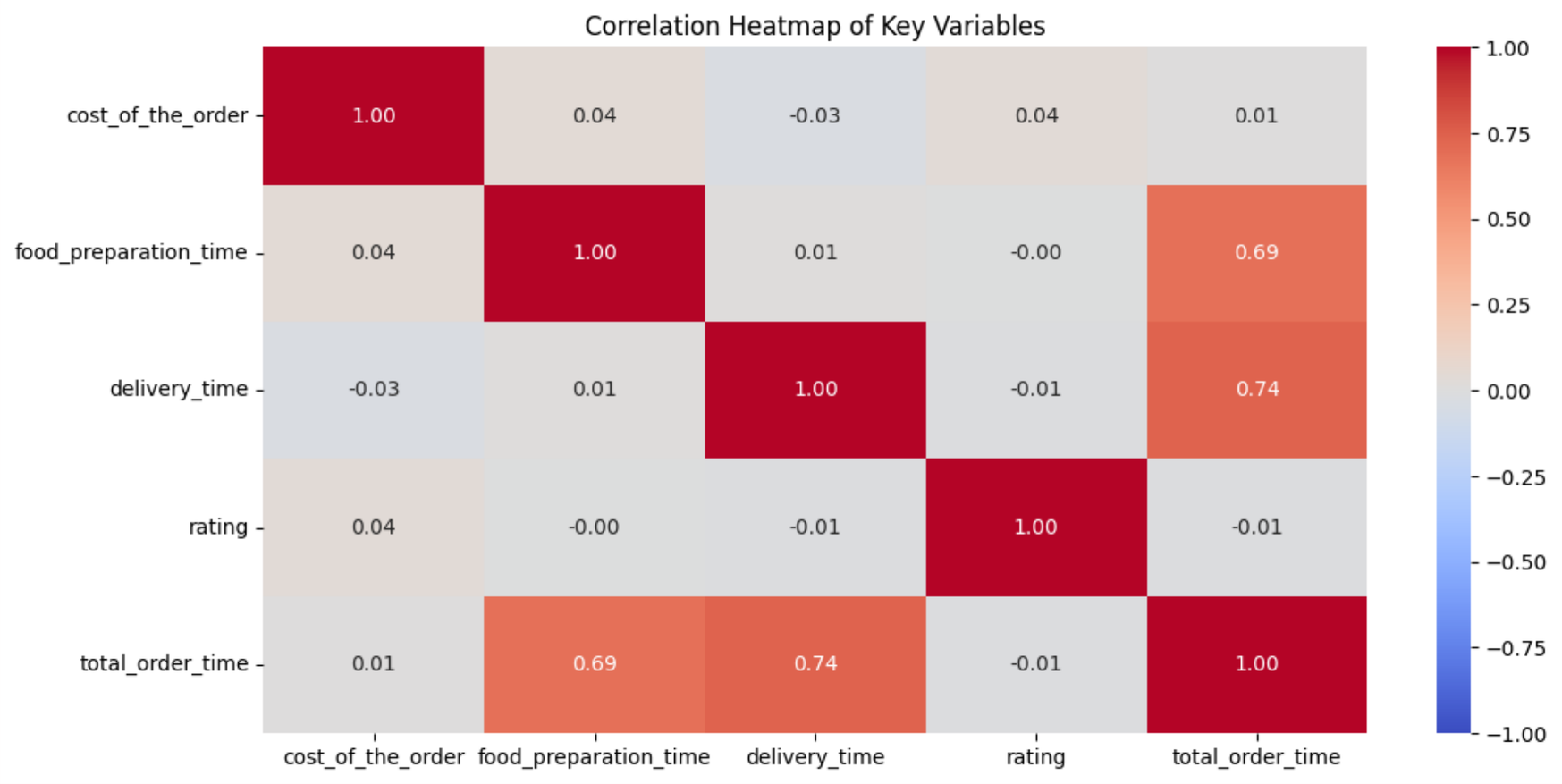
Premium Perception

Orders rated 5 have the highest average cost, suggesting customers may associate higher-cost meals with better quality or experience.

Value Expectations

Lower ratings on expensive orders might indicate unmet expectations when customers pay premium prices.

Correlation Analysis of Key Variables



Total order time is heavily influenced by both delivery time (0.74 correlation) and food preparation time (0.69 correlation). Interestingly, the cost of the order shows very weak correlations with delivery time (-0.03) and food preparation time (0.04), suggesting higher-priced orders don't necessarily take longer to prepare or deliver. Customer ratings have values close to zero when compared to all other variables, indicating that food quality perception is not directly influenced by cost, delivery time, or preparation time.

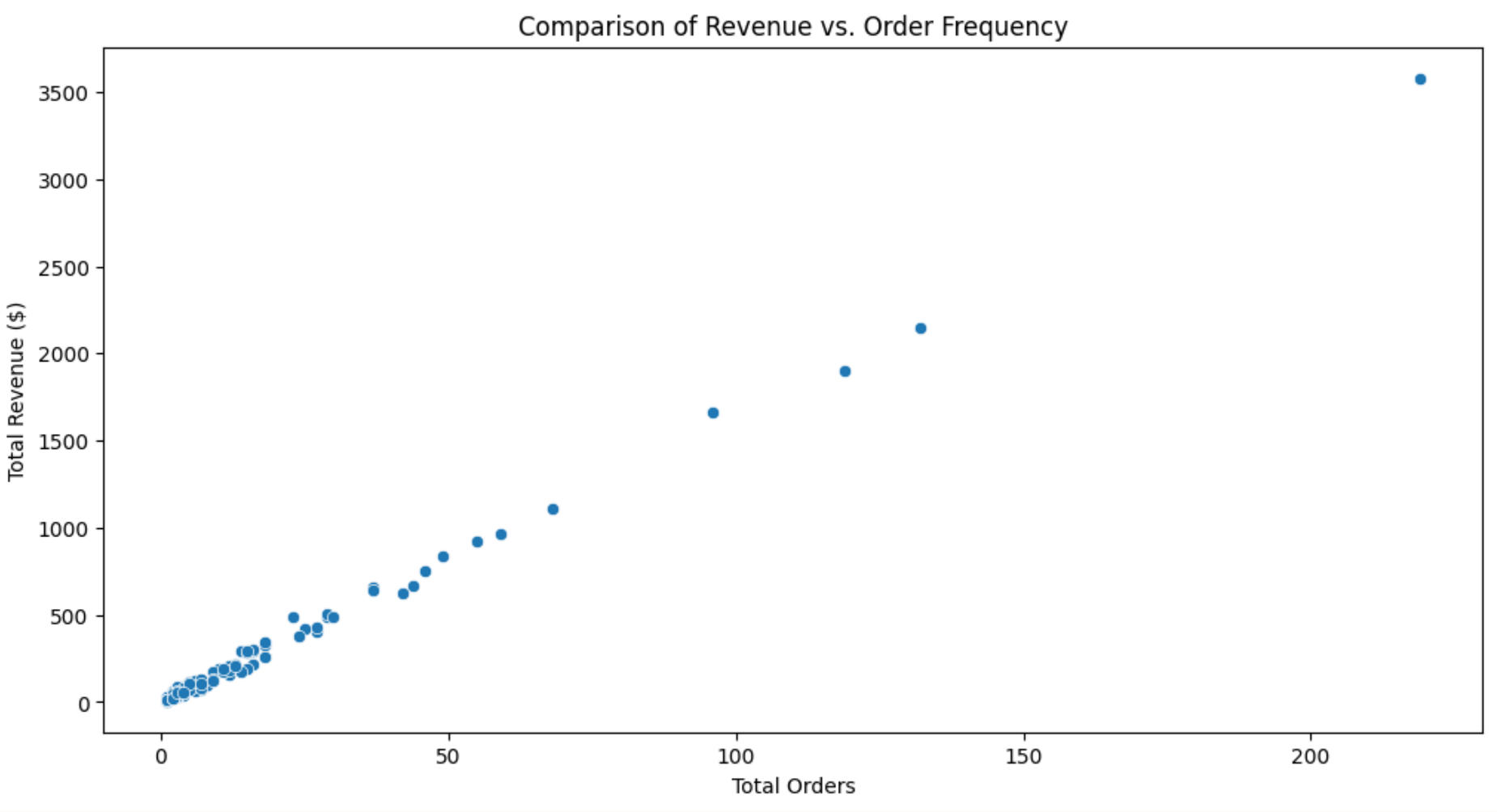
Observations from the Scatter Plot (Revenue vs. Order Frequency): Positive Correlation Between Total

Orders and Revenue:

- The scatter plot shows an upward trend, indicating that restaurants with more orders generally generate more revenue.
- This confirms that order frequency is a strong driver of total revenue.

Outlier Restaurant (Shake Shack):

- One restaurant significantly stands out with the highest order count and revenue.
- This suggests that this restaurant is succeeding due to both high volume and high revenue, making it a top performer.



Cluster of Low-Order, Low-Revenue Restaurants:

- Many restaurants are clustered in the bottom left, meaning they have both low order volume and low revenue.
- These restaurants may not be as competitive or have lower demand.

Question 13:

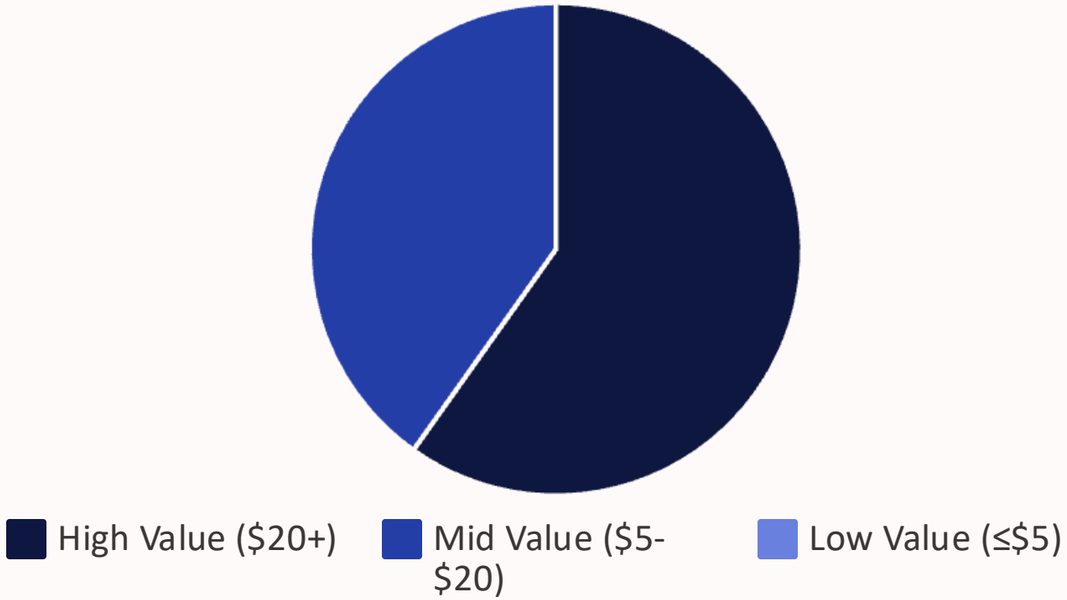
Promotional Offer Qualification Analysis

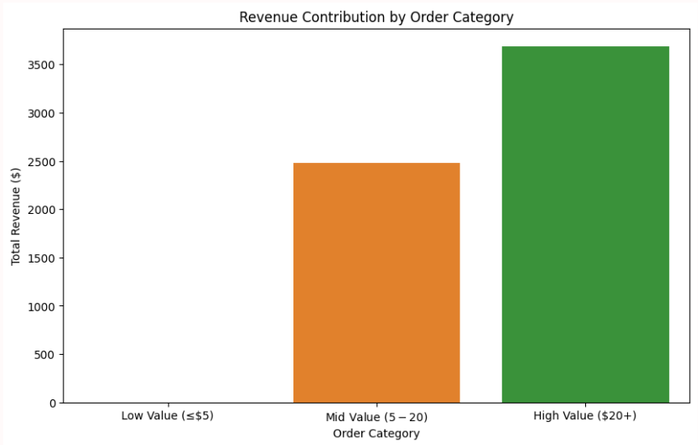
Restaurant Name	Average Rating
The Meatball Shop	4.33
Blue Ribbon Fried Chicken	4.22
RedFarm Broadway	4.17
Shake Shack	4.17
Blue Ribbon Sushi	4.13
RedFarm Hudson	4.11
Parm	4.07

The company wants to provide a promotional offer to restaurants with more than 50 ratings and an average rating greater than 4. Seven restaurants qualify for this promotion, with The Meatball Shop having the highest average rating (4.33). Many of these restaurants were also identified as top revenue generators, suggesting that high revenue is linked to both high volume and strong customer satisfaction.

Question 14

Company Revenue Analysis





Revenue Structure

The company charges restaurants 25% on orders costing more than $20 and 15% on orders costing more than $5. Orders below $5 do not generate any revenue. The total revenue generated by the company across all orders is $6,166.30.

High-value orders ($20+) contribute the most to the company's earnings at $3,688.73 (59.8%), while mid-value orders ($5-$20) generate $2,477.58 (40.2%). This suggests that encouraging larger orders would significantly increase company revenue.

Question 15:

Orders Exceeding 60-Minute Delivery

Percentage of Delayed Orders

Only 10.54% of total orders took more than 60 minutes from order placement to delivery. This suggests that most orders are delivered within an hour, which is a good benchmark for service efficiency.

Cuisines with Most Delayed Orders

Japanese and American cuisines have 55 delayed orders each, followed by Italian (34) and Chinese (31). These cuisines might involve complex preparation processes, high order volume, or delivery challenges.

Impact on Customer Satisfaction

Interestingly, orders taking more than 60 minutes and orders delivered within 60 minutes both have an average rating of 4.21. This suggests that delivery time alone does not strongly influence customer satisfaction.

Question 16: The company wants to analyze the delivery time of the orders on weekdays and weekends. How does the mean delivery time vary during weekdays and weekends?

Observations:

- Deliveries Are Faster on Weekends (22.47 min) Compared to Weekdays (28.34 min)
- Orders placed on weekends are delivered about 6 minutes faster on average than weekday orders.
- This suggests that delivery operations might be more efficient on weekends, possibly due to lower traffic congestion or optimized staffing.