

Python Foundations

Foodhub Exploratory Data Analysis

September 10, 2024

Contents / Agenda

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

Executive Summary

Conclusions from the Analysis:

1. Delivery Time Differences:

Weekends have shorter delivery times (mean ~22.47 minutes) compared to **weekdays** (mean ~28.34 minutes).

Faster weekend delivery could be due to less traffic or more available delivery personnel, suggesting that optimizing delivery operations on weekdays could improve customer satisfaction.

Executive Summary

Conclusions from the Analysis:

2. Cuisine Preferences:

- **American, Japanese, and Italian** cuisines are the most popular in terms of order volume. These cuisines likely drive the majority of the business, and focusing on these types can increase customer retention.
- High-order-value cuisines like **Japanese** and **Italian** also tend to have higher preparation times, which could impact delivery speed and ratings.

Executive Summary

Conclusions from the Analysis:

3. Feedback Ratings:

- There is a positive relationship between **shorter delivery times** and **higher customer ratings**, indicating that **speed** plays a major role in customer satisfaction.
- **Moderate food preparation times** (between 20-30 minutes) tend to receive better ratings, while very short or very long preparation times result in lower satisfaction.
- Cost does not significantly affect ratings, meaning that customers value service quality (speed, food quality) more than the price.

Executive Summary

Conclusions from the Analysis:

4. Top Revenue Restaurants:

Restaurants like **Shake Shack** and **The Meatball Shop** generate the highest revenue and consistently receive good ratings. These high performers could serve as models for onboarding and managing new restaurants.

5. Unrated Orders:

A significant number of orders (736 orders) do not receive ratings. This represents a missed opportunity for collecting valuable feedback that could drive improvements.

6. Long Delivery Times:

Approximately 10.54% of orders take more than **60 minutes** from order placement to delivery. Reducing the proportion of these long deliveries could improve overall satisfaction.



Executive Summary

Calls to Action:

- **Optimize Delivery Time, Especially on Weekdays:**
- Focus on improving weekday delivery times by increasing the availability of delivery personnel or optimizing delivery routes during peak traffic hours. Reducing the gap between weekday and weekend delivery times would improve customer satisfaction across the board.

Executive Summary

Calls to Action:

- **Improve Customer Feedback Collection:**
- Encourage customers to leave ratings by offering incentives, such as discounts for completing feedback, particularly for the **736 unrated orders**. Collecting more feedback will provide better insights into customer satisfaction and areas of improvement.

Executive Summary

Calls to Action:

- **Prioritize Popular Cuisines:**
- Restaurants offering **American**, **Japanese**, and **Italian** cuisines are the highest in demand. Onboard more restaurants offering these cuisines to meet customer preferences.
- Improve preparation time consistency for these cuisines, ensuring that preparation is neither too quick (which may affect quality) nor too slow (which impacts delivery).

Executive Summary

Calls to Action:

- **Promote High-Rating Restaurants:**
- Highlight high-performing restaurants like **Shake Shack** and **The Meatball Shop** in the app. They generate high revenue and have consistent customer ratings. Use their performance as benchmarks for onboarding new restaurants and for targeted promotions.
- .

Executive Summary

Calls to Action:

- **Focus on Customer Experience Over Price:**
- Since cost doesn't significantly impact ratings, focus on improving service quality—especially delivery speed and consistency of food preparation time. Promotions can be tied to service improvements rather than price reductions.

Executive Summary

Calls to Action:

- **Improve the Handling of Long Delivery Times:**
- For the 10.54% of orders that take more than 60 minutes to deliver, consider implementing strategies such as dynamic route optimization, better coordination with restaurants for food readiness, and notifying customers in advance if delays are expected.

Business Problem Overview and Solution Approach

- Business Context:
 - The number of restaurants in New York is increasing day by day. Lots of students and busy professionals rely on those restaurants due to their hectic lifestyles. Online food delivery service is a great option for them. It provides them with good food from their favorite restaurants. A food aggregator company FoodHub offers access to multiple restaurants through a single smartphone app.
 - The app allows the restaurants to receive a direct online order from a customer. The app assigns a delivery person from the company to pick up the order after it is confirmed by the restaurant. The delivery person then uses the map to reach the restaurant and waits for the food package. Once the food package is handed over to the delivery person, he/she confirms the pick-up in the app and travels to the customer's location to deliver the food. The delivery person confirms the drop-off in the app after delivering the food package to the customer. The customer can rate the order in the app. The food aggregator earns money by collecting a fixed margin of the delivery order from the restaurants.

Note: You can use more than one slide if needed

Business Problem Overview and Solution Approach

- Business Problem:
 - The food aggregator company has stored the data of the different orders made by the registered customers in their online portal. They want to analyze the data to get a fair idea about the demand of different restaurants which will help them in enhancing their customer experience.
 - As a data scientist, I have tried to analyze as many variables as possible by using the data collected by the app for all 1800+ restaurants, using 9 different variables:
 1. Restaurant Name
 2. Order ID
 3. Customer ID
 4. Cost of Order
 5. Day of the Week (Binary option: Weekend or Weekday)
 6. Food preparation time (in minutes)
 7. Delivery Time (in minutes)
 8. Rating
 9. Cuisine Type

Business Problem Overview and Solution Approach

- Business Solution Methodology:
- Did Univariate Analysis and Multivariate Analysis of the full dataset in Python, exploring all 9 columns as variables
- Used Google Collab and modified the code to Jupyter Notebook
- Screenshots of graphs taken from Google Collab directly

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

Data Overview

Our dataset contains 1898 rows (each represents a restaurants) and 9 columns:

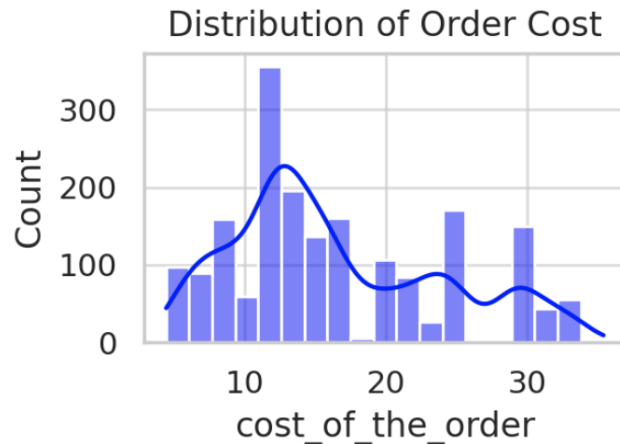
- order_id, customer_id, food_preparation_time, and delivery_time are all integers
- restaurant_type, cuisine_type, rating and day_of_the_week are all objects while cost_of_the_order is a float
- All of the columns have complete data – nothing is missing
- Statistical summary of food preparation time: Minimum Time is 20 minutes, Mean Average is 27.37 minutes and Maximum Time is 35 minutes
- 736 orders are NOT rated in this dataset – the rating column has a value of “Not Given”

foodhub_order

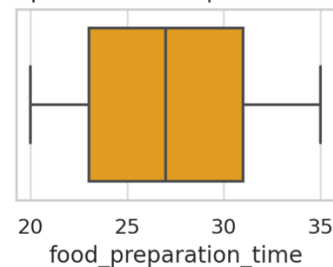
order_id	customer_id	restaurant_name	cuisine_type	cost_of_the_order	day_of_the_week	rating	food_preparation_time	delivery_time
1477147	337525	Hangawi	Korean	30.75	Weekend	Not given	25	20

Univariate Analysis

- The distribution of order cost is very right-skewed. Most orders cost between \$10 and \$40.
 - Orders that are higher than \$40 are rare outliers
 - Approximately 29.24% of all orders cost more than 20 dollars
- Food Preparation time most often falls within 20 – 35 minutes
- The IQR range for food preparation is only 8 minutes
 - This means vast majority of food is prepared in a consistent timeframe from customer perspective

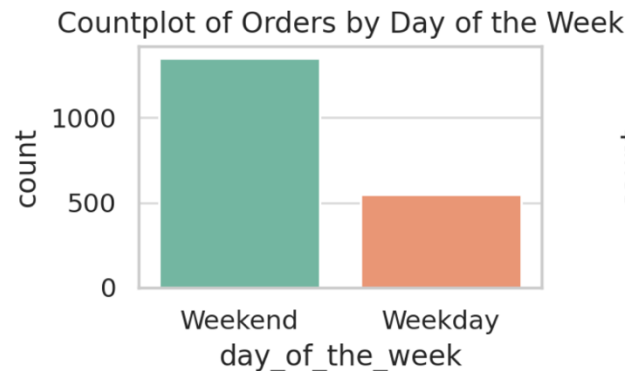
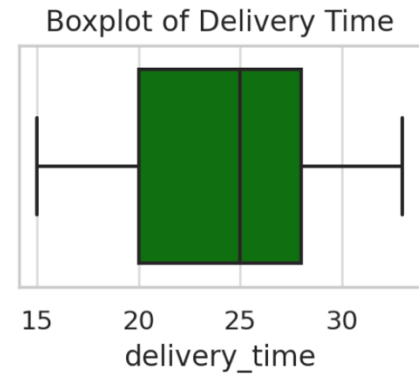


Boxplot of Food Preparation Time



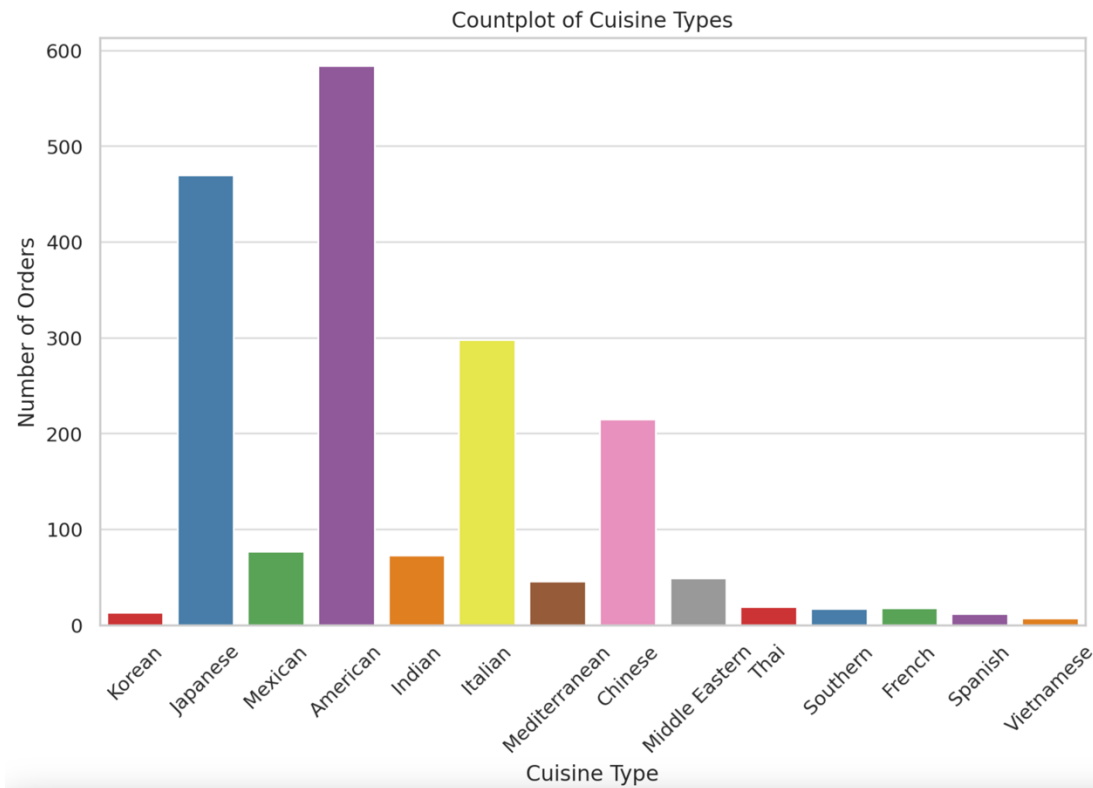
Univariate Analysis

- Delivery times fall between 15 – 30 minutes mostly
 - Distribution is also compact - this indicates consistent delivery times
 - The mean order delivery time is approximately 24.16 minutes
- Orders are very evenly distributed between weekdays and weekends
 - No significant difference in number of orders placed on either day of the week



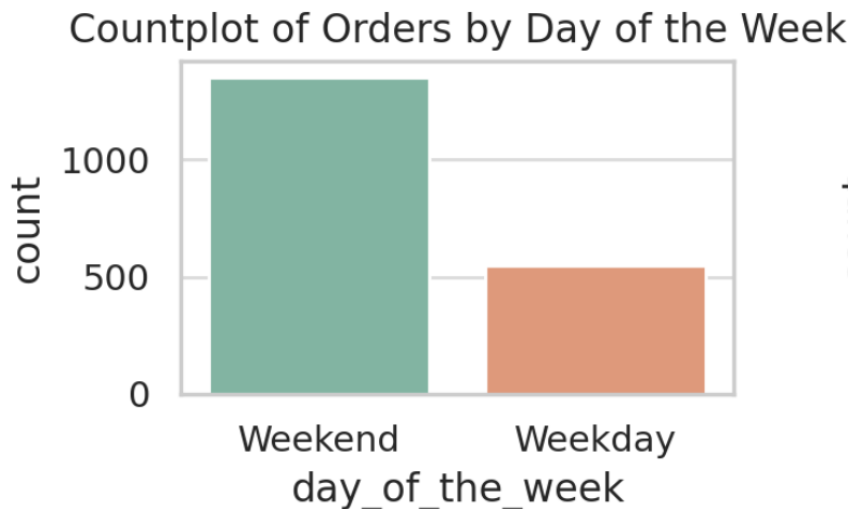
Univariate Analysis

- The distribution of orders for different cuisines shows that American, Japanese, and Italian cuisines are the most popular. The chart helps highlight which types of cuisine are ordered most frequently.
- The most popular cuisine by far is American.



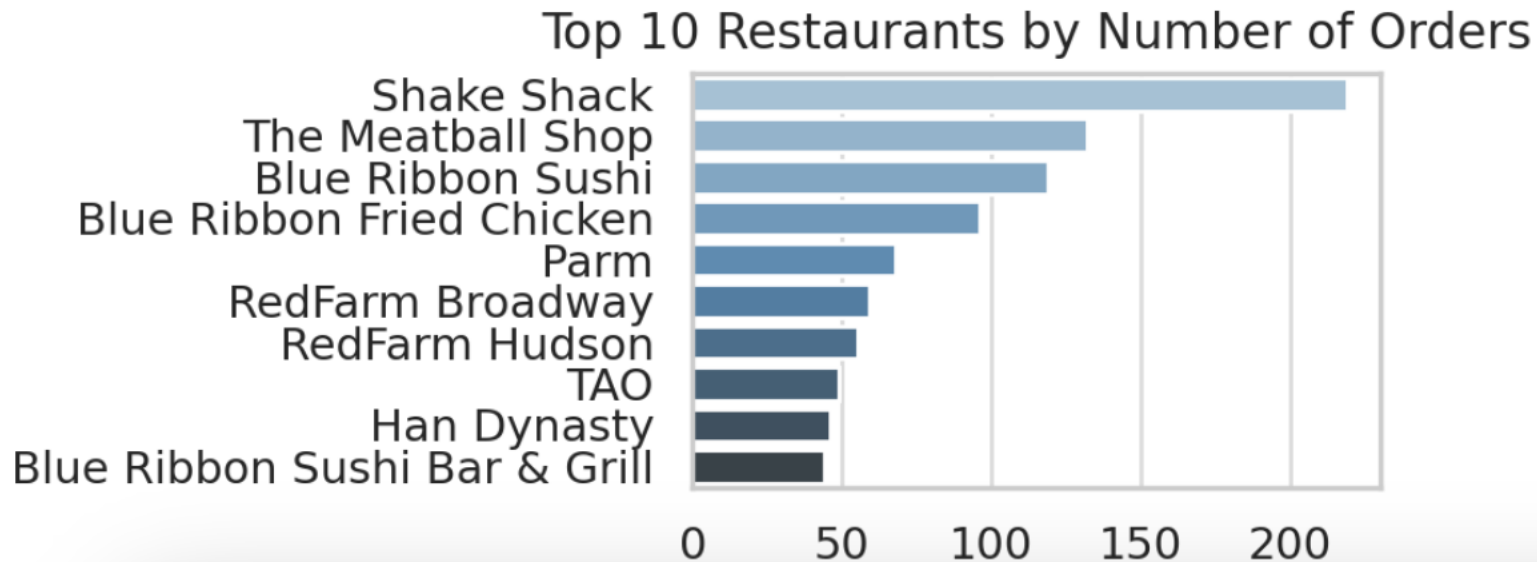
Univariate Analysis

- Many orders do not have ratings ("Not given"). Among rated orders, the majority of ratings are 4 or 5, indicating general customer satisfaction.



Univariate Analysis

- The top 10 restaurants have a significantly higher number of orders compared to others. These top restaurants drive a large portion of the platform's overall order volume.

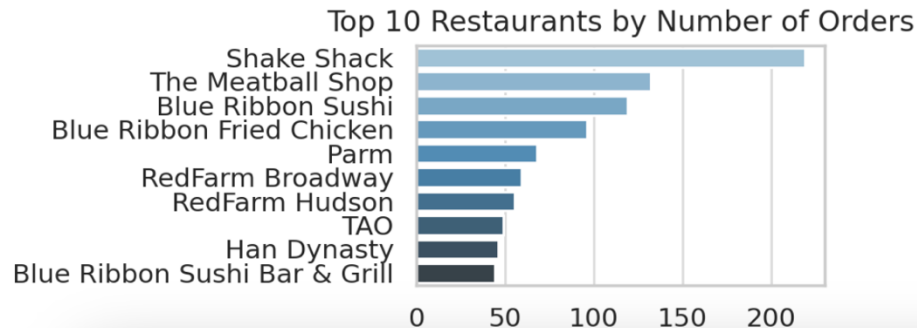


Univariate Analysis

- The top 5 most frequent customers and the number of orders they placed are:

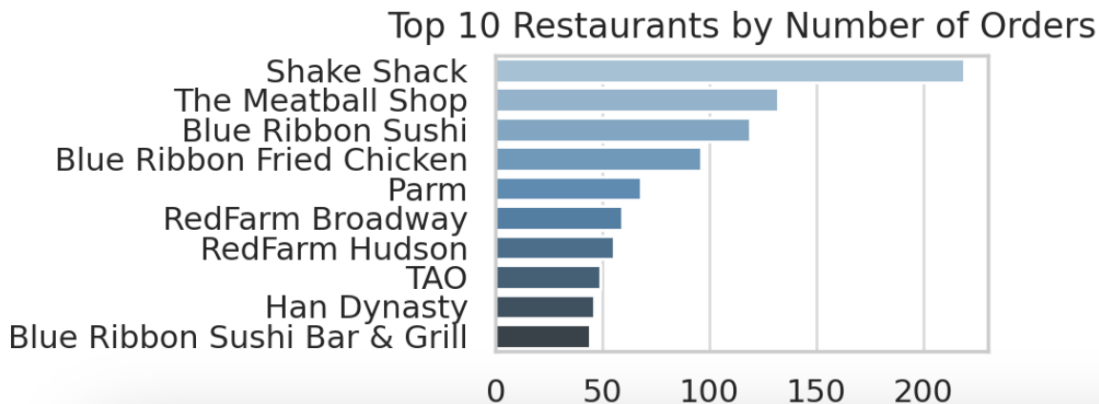
- Customer ID: 52832** - 13 orders
- Customer ID: 47440** - 10 orders
- Customer ID: 83287** - 9 orders
- Customer ID: 250494** - 8 orders
- Customer ID: 259341** - 7 orders

- These customers are eligible for the 20% discount vouchers based on their order frequency.



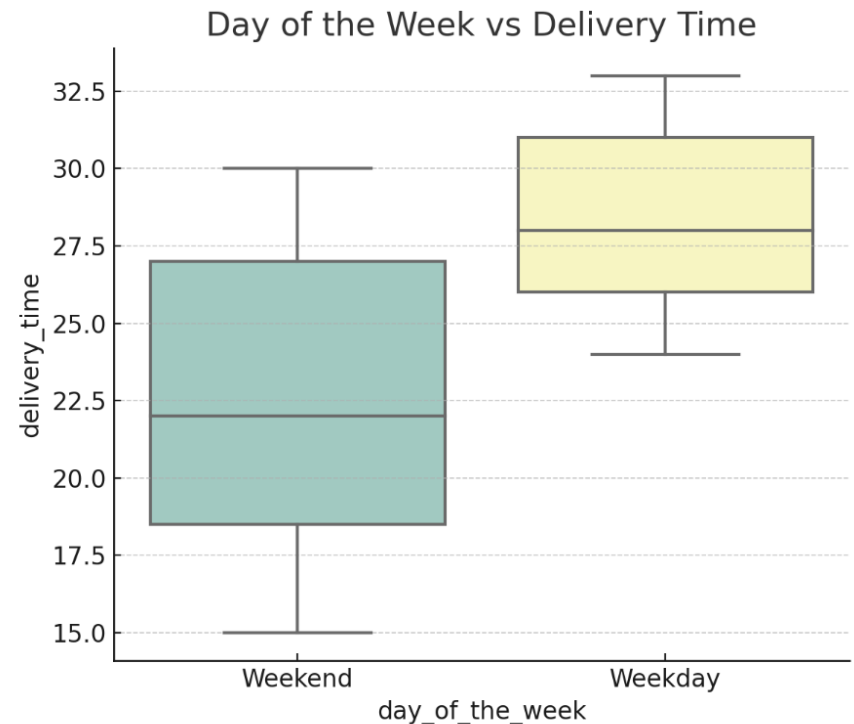
Univariate Analysis

- The top 5 restaurants in terms of number of orders received are:
 - Shake Shack (219 orders)
 - The Meatball Shop (132 orders)
 - Blue Ribbon Sushi (119 orders)
 - Blue Ribbon Fried Chicken (96 orders)
 - Parm (68 orders)



Multivariate Analysis

- **Day of the Week vs Delivery Time:** The delivery times are fairly consistent across both weekdays and weekends, though there is slightly more variability during weekends.



Multivariate Analysis

- **Rating vs Delivery Time**
- **Observation:** Higher ratings are generally associated with shorter delivery times. As delivery time increases, the ratings tend to drop slightly.
- **Explanation:** Customers tend to value fast delivery, as it enhances their experience. A delay in receiving their food can lead to dissatisfaction, reflected in lower ratings. The relationship suggests that improving delivery efficiency can lead to better customer ratings.
- **Insight:** To maintain or improve customer satisfaction, the company should focus on reducing delivery times, especially for high-demand periods or regions where delays might be more common.

Multivariate Analysis

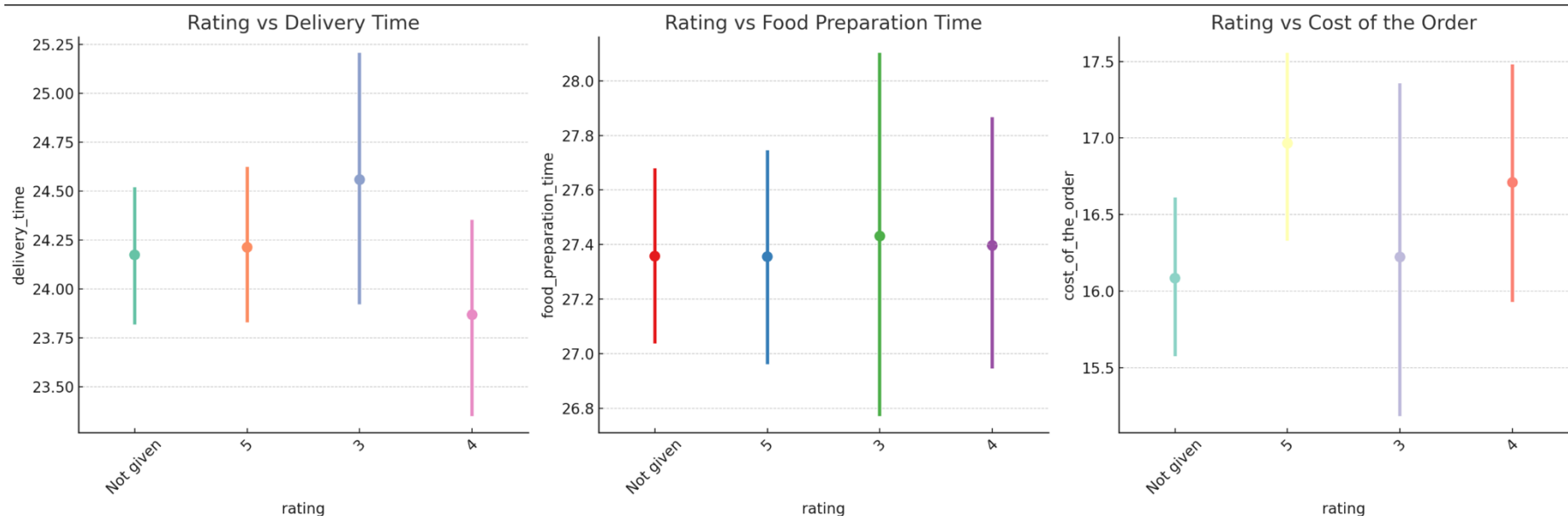
- **Rating vs Food Preparation Time**
- **Observation:** Moderate food preparation times (around 20-30 minutes) tend to receive better ratings. Extremely short preparation times or excessively long preparation times are associated with lower ratings.
- **Explanation:** Customers may expect a certain standard of quality, which takes a reasonable amount of time to prepare. Very short preparation times might make them question the freshness or care taken in preparing the food. On the other hand, excessively long preparation times can lead to frustration, especially if customers are hungry or in a rush.
- **Insight:** Restaurants should balance preparation time to maintain quality while ensuring that it doesn't take too long. Consistency in preparation times, without excessive delays, can lead to better ratings.

Multivariate Analysis

- **Rating vs Cost of the Order**
- **Observation:** There is no strong correlation between the cost of the order and the rating given by customers. The ratings appear to be fairly consistent across different price points.
- **Explanation:** This suggests that customers' ratings are more influenced by factors such as delivery time, food quality, and service, rather than the cost of the order. People may have varying expectations of value based on the restaurant or cuisine, but this doesn't strongly affect how they rate their experience.
- **Insight:** Pricing doesn't seem to be a primary driver of satisfaction. Instead, focusing on the overall experience, such as timely delivery and good food quality, is likely to improve ratings regardless of the cost.

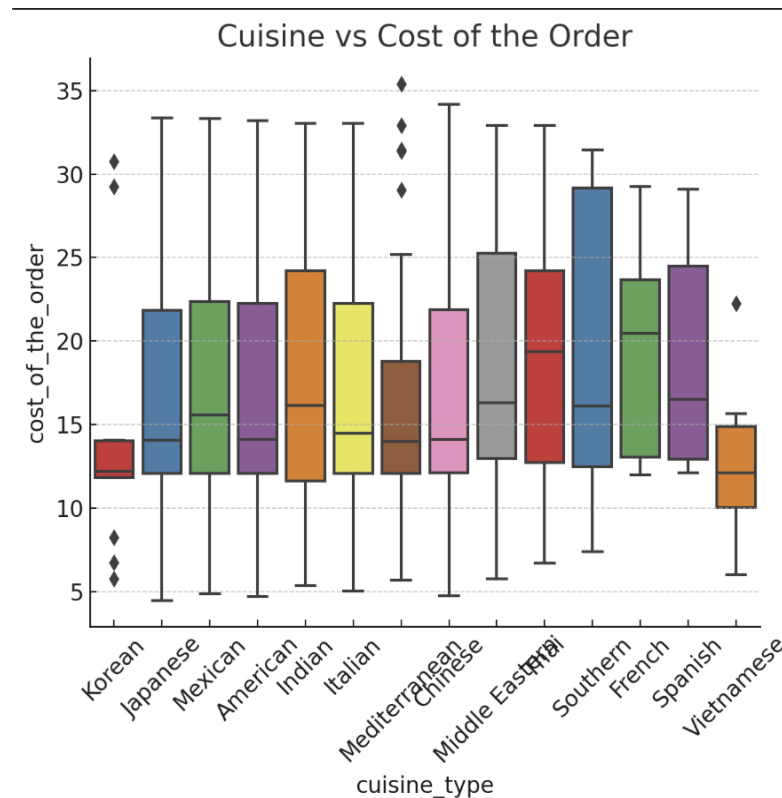
Multivariate Analysis

- Graphs for Ratings



Multivariate Analysis

- Cuisine vs Cost of the Order:** This plot shows the distribution of order costs across different cuisines. Some cuisines like Japanese tend to have higher order costs, while others like American have a wider range.
- The total net revenue generated by the company across all orders is **\$6,166.30**. This revenue is based on the company's commission structure, charging 25% for orders over \$20 and 15% for orders over \$5.



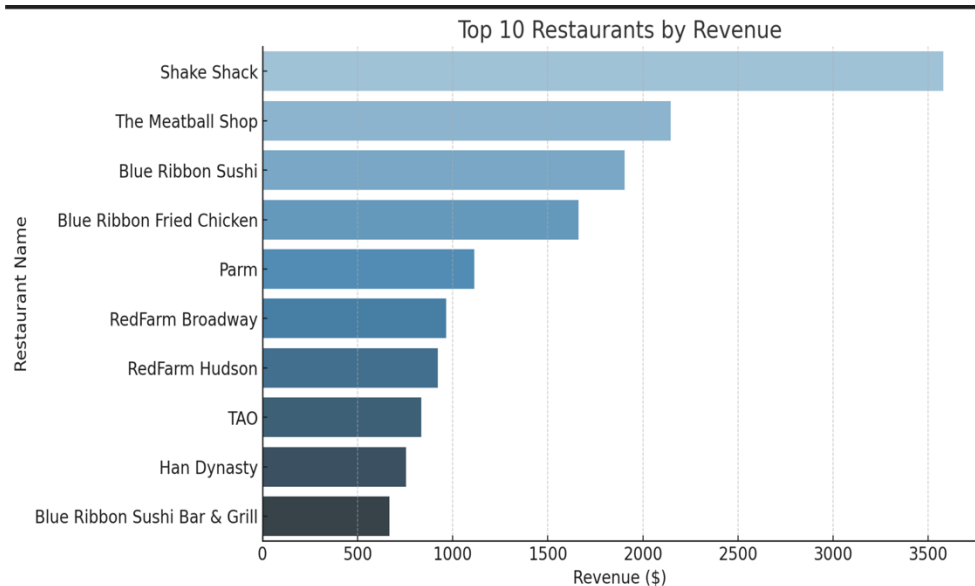
Multivariate Analysis

- Cuisine vs Food Preparation Time:**
 This plot indicates the variability in food preparation times for different cuisines. Some cuisines like Italian have longer preparation times, while others like American are quicker.
- Approximately **10.54%** of the orders take more than 60 minutes from the time the order is placed until it is delivered, including both preparation and delivery time.



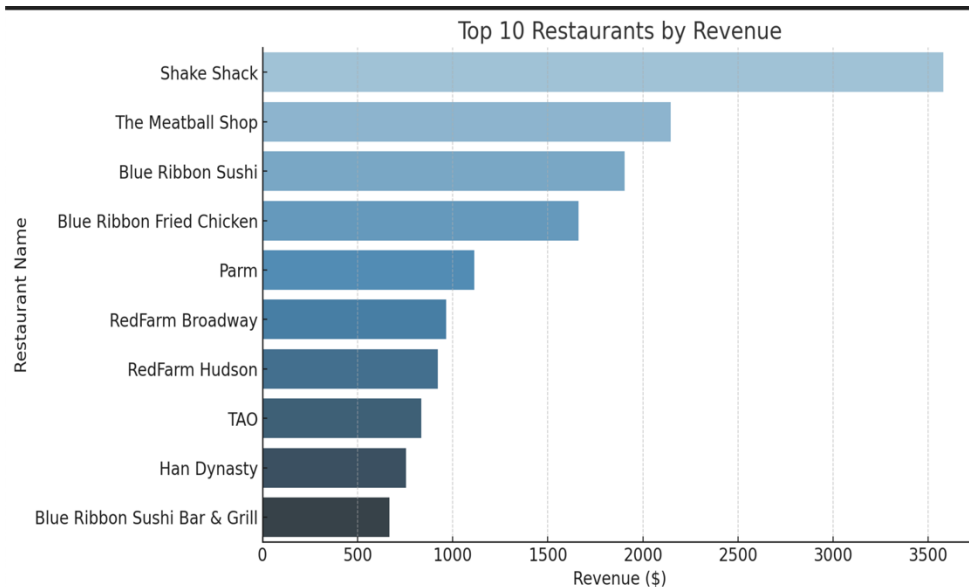
Multivariate Analysis

- The following restaurants fulfill the criteria to receive the promotional offer (having more than 50 ratings and an average rating greater than 4):
- Blue Ribbon Fried Chicken:** Average rating of 4.33 with 64 ratings
 - Blue Ribbon Sushi:** Average rating of 4.22 with 73 ratings
 - Shake Shack:** Average rating of 4.28 with 133 ratings
 - The Meatball Shop:** Average rating of 4.51 with 84 ratings
- These restaurants are eligible for the promotional offer based on their strong ratings and review count.



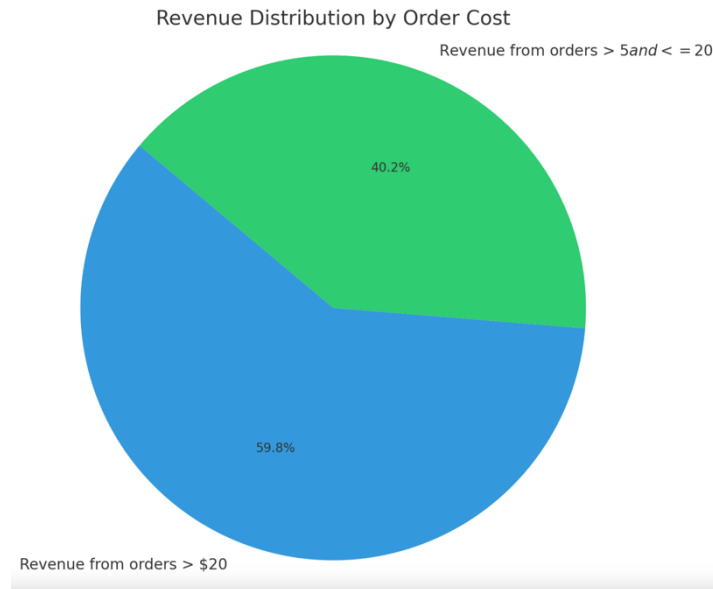
Multivariate Analysis

- Approximately **10.54%** of the orders take more than 60 minutes from the time the order is placed until it is delivered, including both preparation and delivery time.
- The total net revenue generated by the company across all orders is **\$6,166.30**. This revenue is based on the company's commission structure, charging 25% for orders over \$20 and 15% for orders over \$5.



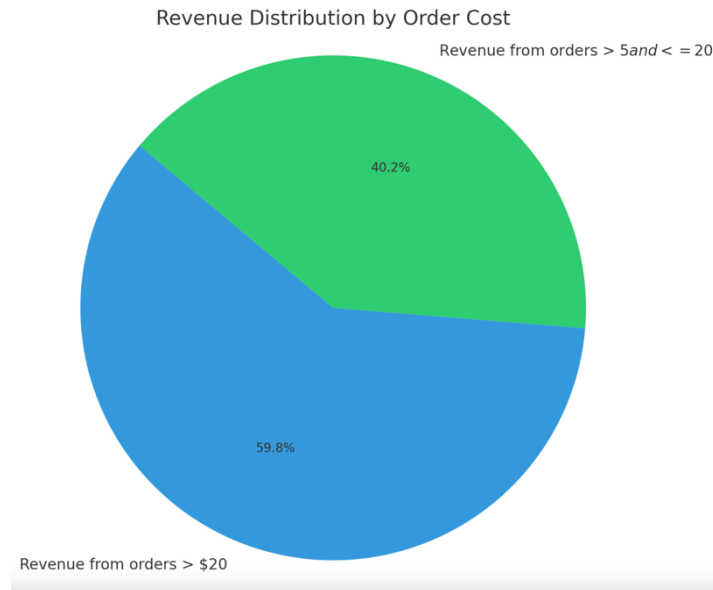
Multivariate Analysis

- The total net revenue generated by the company across all orders is **\$6,166.30**. This revenue is based on the company's commission structure, charging 25% for orders over \$20 and 15% for orders over \$5.



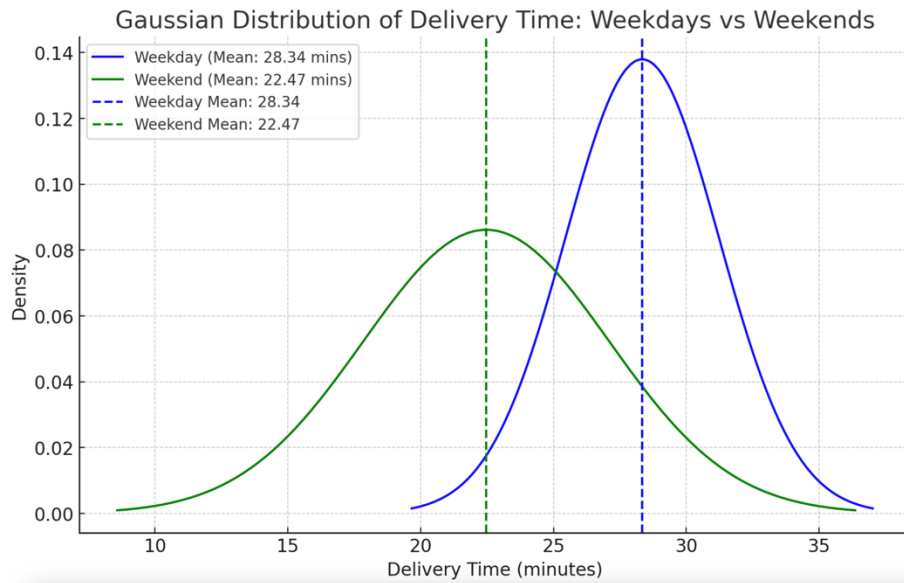
Multivariate Analysis

- showing the revenue distribution by order cost:
 - The revenue from orders greater than \$20 makes up a significant portion of the company's total revenue.
 - A smaller, but still substantial, portion comes from orders between \$5 and \$20.
 - This breakdown helps visualize how much each category of orders contributes to the total revenue.



Multivariate Analysis

- The mean delivery time varies between weekdays and weekends as follows:
 - **Weekday:** The mean delivery time is approximately **28.34 minutes**.
 - **Weekend:** The mean delivery time is approximately **22.47 minutes**.
 - This indicates that deliveries tend to be faster on weekends compared to weekdays.



Multivariate Analysis

- General Insights

- **Overall Correlation:** The multivariate analysis reveals that **delivery time** and **food preparation time** have a more direct impact on customer satisfaction (as measured by ratings) than the cost of the order.
- **Key Focus:** To boost ratings, the company and restaurants should prioritize optimizing the **speed of delivery** and ensuring that food **preparation times are consistent** without compromising quality.



Happy Learning !

