Data Analysis - Low Code

Project Python Foundations
FoodHub

October 2, 2025

Andres Dodero

Contents / Agenda

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

Executive Summary

- **Key Conclusions**
  - **Operational Metrics**: Cost, preparation time, and delivery time show minimal correlation, weekday vs. weekend show different delivery times.
  - **Customer Satisfaction**: Feedback ratings are weakly correlated with cost and preparation time, and almost unrelated to delivery time. This suggests food quality outweigh delivery speed.
  - **Cuisine Performance**:
    - **Rating**: Spanish, Thai and Indian have the highest rating.
    - **Preparation Times**: Korean and Vietnamese have the lowest times; Vietnamese and Italian have the longest times, when looking at Mean
  - **Revenue Opportunities**: increase demand for higher cost types of cuisine, promote restaurants with lower preparation times and improve delivery efficiency/speed for higher productivity.

Executive Summary

- **Actionable Insights**
  - Speed is not the main driver of satisfaction; focus on food quality.
  - Cuisine type significantly influences ratings; leverage high-rated cuisines for growth and high-cost for revenue.
  - Operational independence of cost, prep time, and delivery time allows targeted optimization without trade-offs.

Executive Summary

- **Recommendations:**
  - **Promote high-cost and high-rated cuisines:** (Indian, Thai, Spanish) and improve low-rated ones (Vietnamese, Korean, Mediterranean), add more restaurants on high-rated/cost category
  - **Promote low preparation time cuisine or restaurants** particularly on weekdays where delivery time is longer.
  - **Optimize delivery logistics** with predictive models for traffic and distance.
  - **Streamline communication with high prep time restaurants** to look for improvements, share data (healthy competition) and increase efficiency

Business Problem Overview and Solution Approach

- ## Problem Statement:
  - **FoodHub**, a food aggregator app in New York, is experiencing rapid growth in restaurant partnerships due to increasing demand from students and busy professionals. The app facilitates online food orders, assigns delivery personnel, and collects customer ratings. FoodHub earns revenue by taking a fixed margin from each delivery order. The company needs to explore the order dataset and answer key business questions that will help FoodHub make data-driven decisions.
  - Need to Analyze its order data to:
    - Understand customer demand across different restaurants.
    - Identify performance trends (e.g., ratings, preparation and delivery times).
    - Improve customer experience and optimize business operations.

Business Problem Overview and Solution Approach

- Solution approach / methodology:
  - **Main Goal:** uncover insights about restaurant demand, customer behavior, and operational efficiency.
  - **Data Preparation & Exploration:** Load and clean the dataset, check for missing values, and perform descriptive statistics to understand the structure and distribution of key variables.
  - **Univariate & Multivariate Analysis:** Use visualizations (histograms, boxplots, countplots) to explore individual variables & relationships between features like cuisine type, cost, rating, and delivery time.
  - **Business Question Resolution:** Answer targeted questions on top restaurants, customer behavior, popular cuisines, delivery efficiency, and promotional eligibility using filtering, grouping, and aggregation techniques.
  - **Revenue & Operational Insights:** Calculate net revenue using a tiered margin model, analyzed total delivery time, and compared performance across weekdays and weekends to inform business strategy.

Data Overview

- **Question 1:** How many rows and columns are present in the data? There are 1,898 rows and 9 columns
- **Question 2:** What are the datatypes of the different columns in the dataset?

Column	Description	Data Type
order_id	Unique identifier for each order	Integer / String
customer_id	Unique identifier for the customer	Integer / String
restaurant_name	Name of the restaurant	String
cuisine_type	Cuisine category (e.g., Mexican, Italian)	String
cost_of_the_order	Cost of the order in currency	Float
day_of_the_week	Weekday or weekend indicator	String
rating	Customer rating (1–5) or "Not Given"	Float / String
food_preparation_time	Time (minutes) to prepare food	Integer
delivery_time	Time (minutes) to deliver food	Integer

Data Overview

- **Question 3:** Are there any missing values in the data? If yes, treat them using an appropriate method.
    - Yes, the dataset contains missing values in the rating column. Out of 1,898 total orders, 736 orders have the value "Not Given" for feedback rating.
    - This represents approximately 38.8% of the data for that column.
- **How I treated missing values**
    - I converted "Not Given" to NaN for consistency.
    - For quantitative analysis (such as calculating averages, correlations, and profitability scores), I excluded these rows
- **Why this method is appropriate**
    - Excluding missing ratings avoids misleading conclusions about customer satisfaction.
    - It maintains data integrity for analyses where ratings are critical (e.g., identifying top cuisines by rating).
    - For other metrics (cost, preparation time, delivery time), all rows were retained since those columns had no missing values.

Data Overview

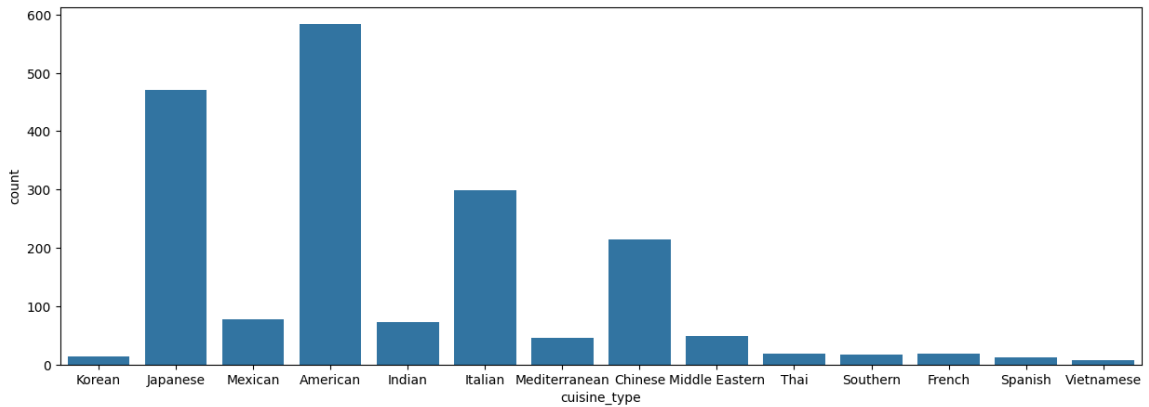
- **Question 4:** What is the minimum, average, and maximum time it takes for food to be prepared once an order is placed?
    - Minimum: 20.00 minutes
    - Average: 27.37 minutes
    - Maximum: 35.00 minutes
- **Question 5:** How many orders are not rated?
    - 736 orders do not have a rating

Univariate Analysis

- Below is a high-level univariate analysis for all columns:
  - **order_id & customer_id:** Both are unique identifiers. order_id is unique for each order, while customer_id repeats for frequent customers.
  - **restaurant_name:** Categorical variable. A few restaurants dominate the order volume (e.g., Shake Shack, Blue Ribbon Sushi).
  - **cuisine_type:** Limited categories. American and Japanese cuisines are the most frequent, followed by Italian and Mexican.
  - **cost_of_the_order:** Numerical value; can be used for financial analysis and will need to round to only 2 decimals.
  - **day_of_the_week:** Two categories: Weekday and Weekend.
  - **Rating:** Includes numeric ratings (1–5) and “Not given.
  - **food_preparation_time:** Numerical (in minutes); will need to adjust the format for any questions asking for number of seconds.
  - **delivery_time:** Numerical, similar to preparation time.

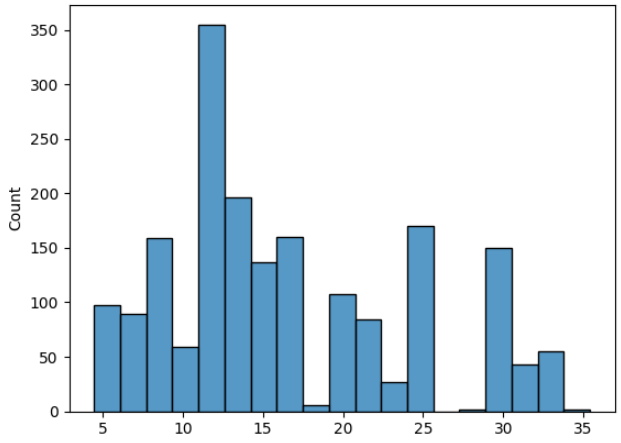
Univariate Analysis

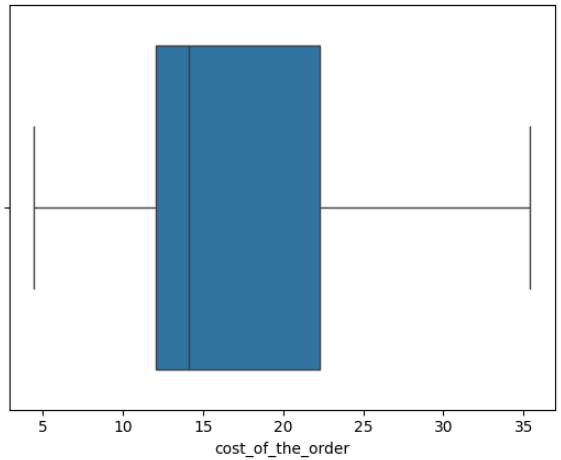
- **Question 6:** All the variables and observations on their distributions
  - 1,898 different orders are included for analysis
  - There are 1,200 unique Customer IDs
  - There are 178 unique Restaurant Names
  - There are 14 Cuisine Types



Univariate Analysis

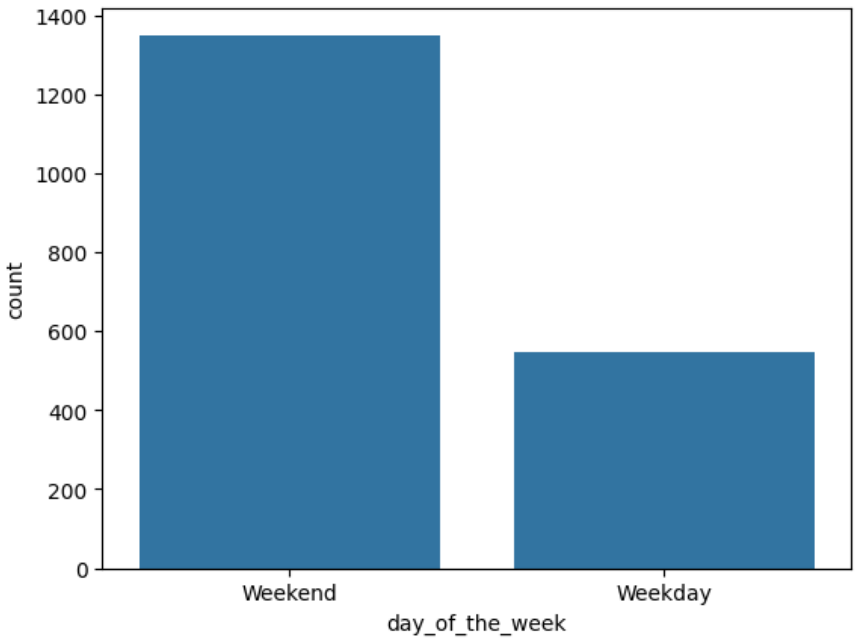
- **Question 6:** Cost of the Order
  - Distribution is right-skewed, where most orders are on the lower end of the cost range
  - Minimum order is $5; maximum: $35
  - Median is about $14.50
  - No extreme outliers beyond the whiskers





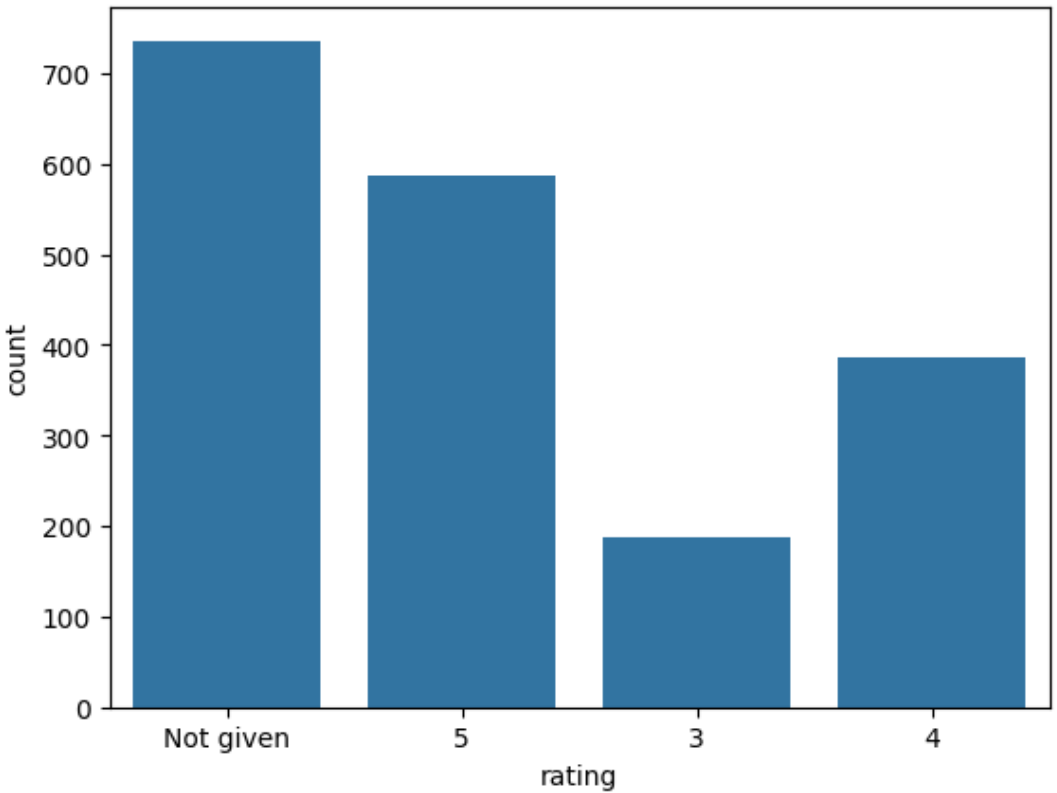
Univariate Analysis

- **Question 6:** Day of the Week
  - Most orders happen on the Weekend (more than twice in volume)
  - Only 2 variables: Weekend and Weekday



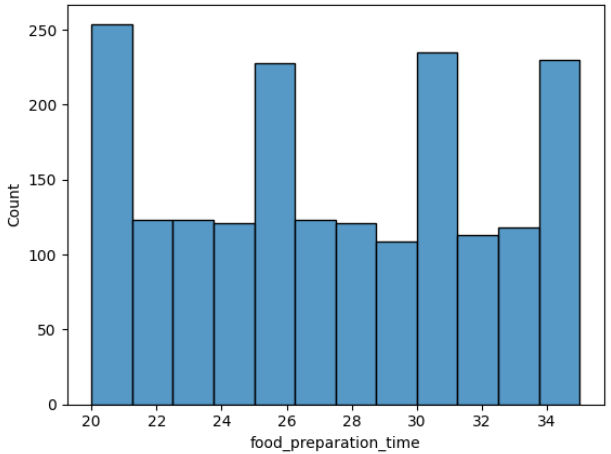
Univariate Analysis

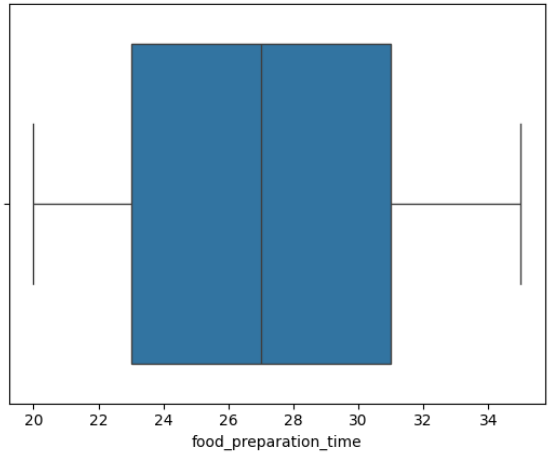
- **Question 6:** Rating
  - Most customers did not rate the food/service
  - After that, most gave the highest rating of 5
  - The lowest rating provided was a 3 of 5
  - Need to encourage more customers to provide ratings.
  - High rating volume suggests strong satisfaction.



Univariate Analysis

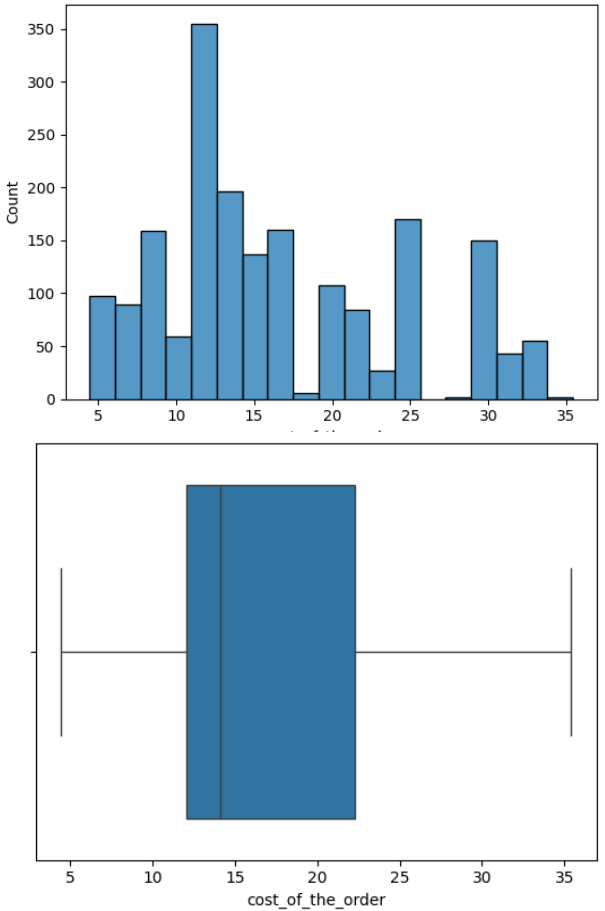
- **Question 6:** Food Preparation Time
  - Fairly uniform distribution between 20-35 min, this shows relative consistent preparation times across restaurants.
  - Median is ~27.5 minutes
  - Whiskers extend from 20 – 35 minutes
  - No extreme outliers





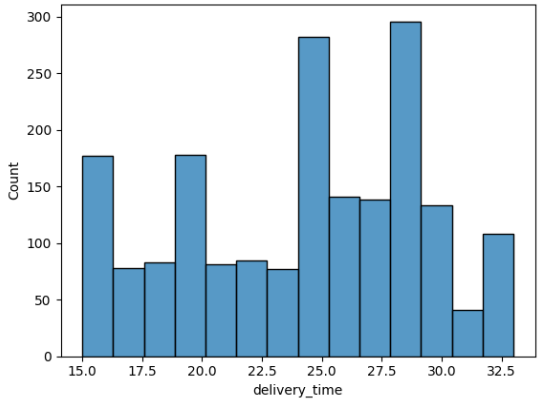
Univariate Analysis

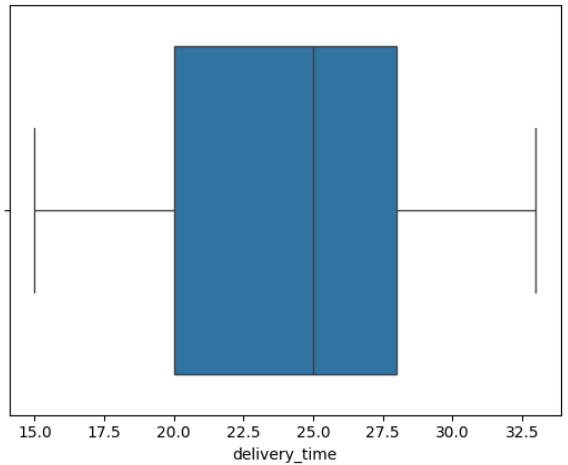
- **Question 6:** Cost of the Order
  - High frequency on $10-$12
  - Min: $5 / Max: $35
  - A few orders are > $30, but are rare
  - Median is ~$14



Univariate Analysis

- **Question 6:** Delivery Time
  - Distribution does not show a strong skew; does have high peaks on 25 and 28 minute delivery times
  - Very few orders exceeded 32 min
  - Whiskers extend from 15 to 33 min
  - Median is 25 minutes





Univariate Analysis

- **Question 7:** Which are the top 5 restaurants in terms of the number of orders received?
  - They are Shake Shack(219), The Meatball Shop (132), Blue Ribbon Sushi (119), Blue Ribbon Fried Chicken (96)and Parm (68)
- **Question 8:** Which is the most popular cuisine on weekends?
  - Cuisine type: American
- **Question 9:** What percentage of the orders cost more than 20 dollars?
  - Percentage of orders above $20: 29.24%
- **Question 10:** What is the mean order delivery time?
  - Mean order delivery time is: 24.16 minutes

Univariate Analysis

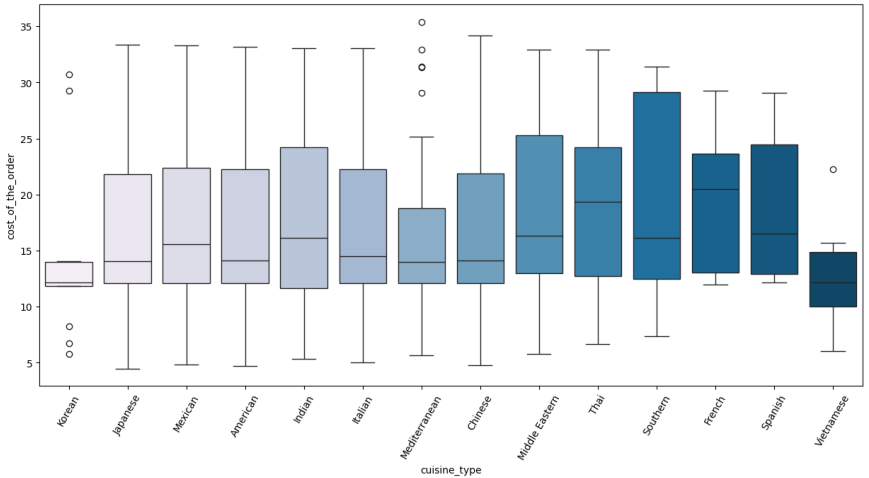
- **Question 11:** The company has decided to give 20% discount vouchers to the top 5 most frequent customers.
    - Find the IDs of these customers and the number of orders they placed.

customer_id	count
52832	13
47440	10
83287	9
250494	8
259341	7

Multivariate Analysis

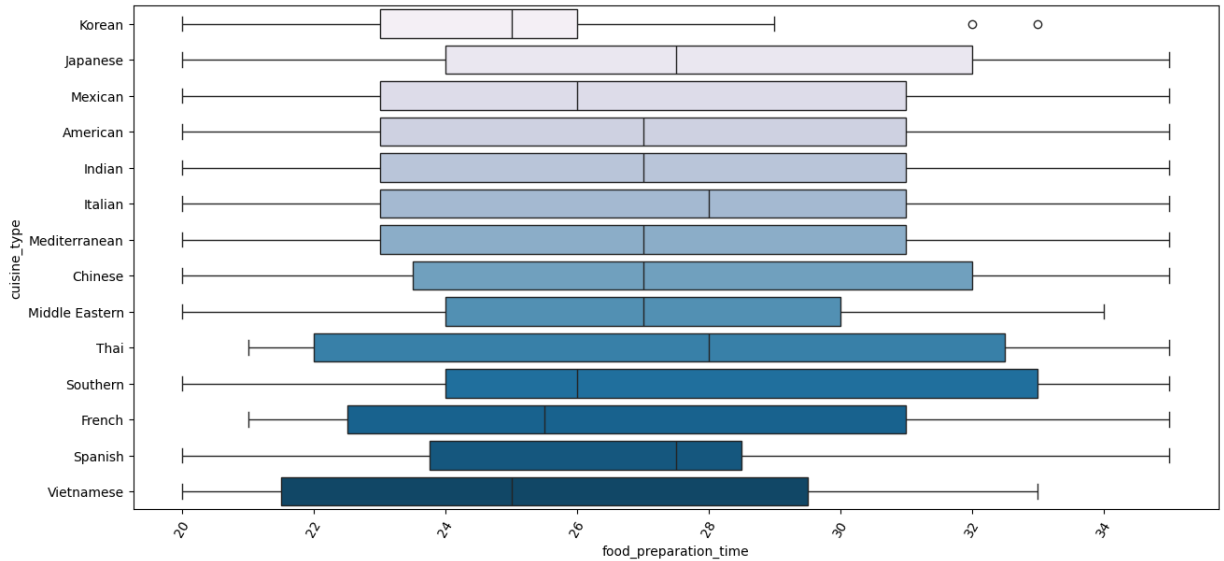
- **Question 12:** Perform a multivariate analysis to explore relationships between the important variables in the dataset. Cuisine vs. Cost of the Order

  - Korean & Vietnamese cuisines have the lowest median
  - French has the highest median
  - Middle East, Indian, Thai and Southern show the widest IQR, meaning prices vary significantly
  - Middle Eastern, Korean and Vietnamese have outliers



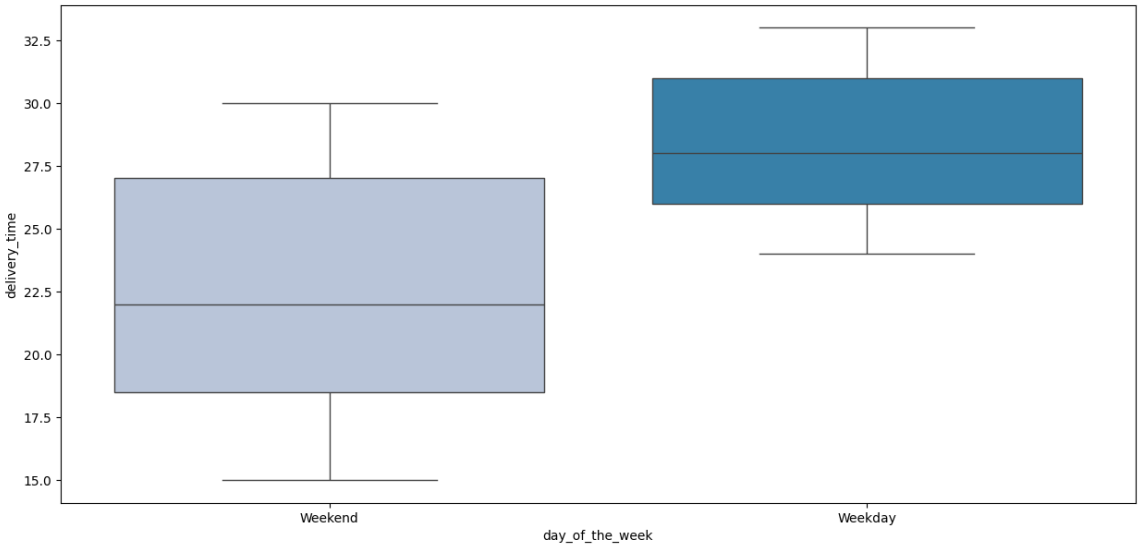
Multivariate Analysis

- **Question 12:** Cuisine vs. Food Preparation Time
  - Korean and Vietnamese have the lowest median times near 25 min
  - Highest median prep time is for Italian and Thai cuisine



Multivariate Analysis

- **Question 12:** Day of the Week vs. Delivery Time
    - Weekdays have a higher delivery time, probably due to traffic
    - No significant outliers



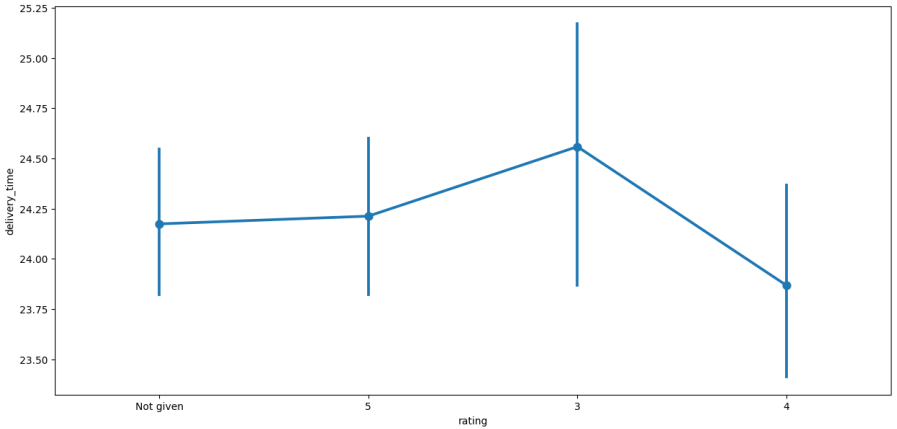
Multivariate Analysis

- **Question 12**: Revenue Generated by the restaurants
    - Shake Shack is the top restaurant generating revenue (cost of order)
    - Recommend for the owner to meet with the top 3-5 restaurant owners for rapport, survey, etc.

restaurant_name	cost_of_the_order
Shake Shack	3579.53
The Meatball Shop	2145.21
Blue Ribbon Sushi	1903.95
Blue Ribbon Fried Chicken	1662.29
Parm	1112.76
RedFarm Broadway	965.13
RedFarm Hudson	921.21
TAO	834.50
Han Dynasty	755.29
Blue Ribbon Sushi Bar & Grill	666.62
Rubirosa	660.45
Sushi of Gari 46	640.87
Nobu Next Door	623.67
Five Guys Burgers and Fries	506.47

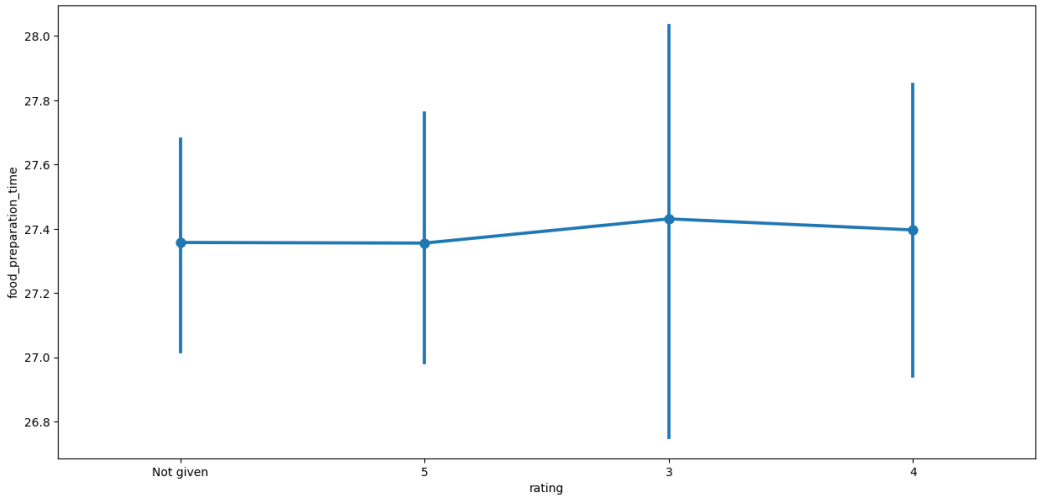
Multivariate Analysis

- **Question 12**: Rating vs. Delivery Time
  - Orders where the delivery time was the highest, had the lowest rating provided (3 out of 5), but the difference is small
  - Delivery time does not strongly influence ratings, but consistently faster deliveries could help maintain high satisfaction.
  - Other factors (food quality, accuracy, etc.) seem to play a bigger role in ratings.



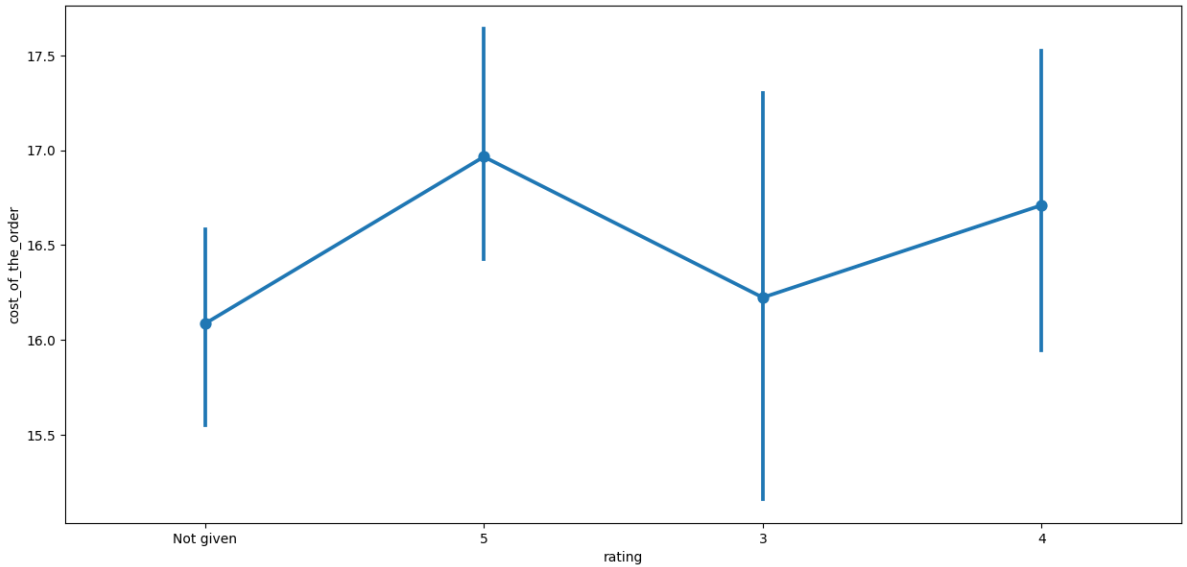
Multivariate Analysis

- **Question 12**: Rating vs. Food Preparation Time
  - Food preparation time seems very consistent
  - Ratings of 3 (lowest) have a slightly higher time, but is minimal
  - Food preparation time does not appear to influence ratings significantly.



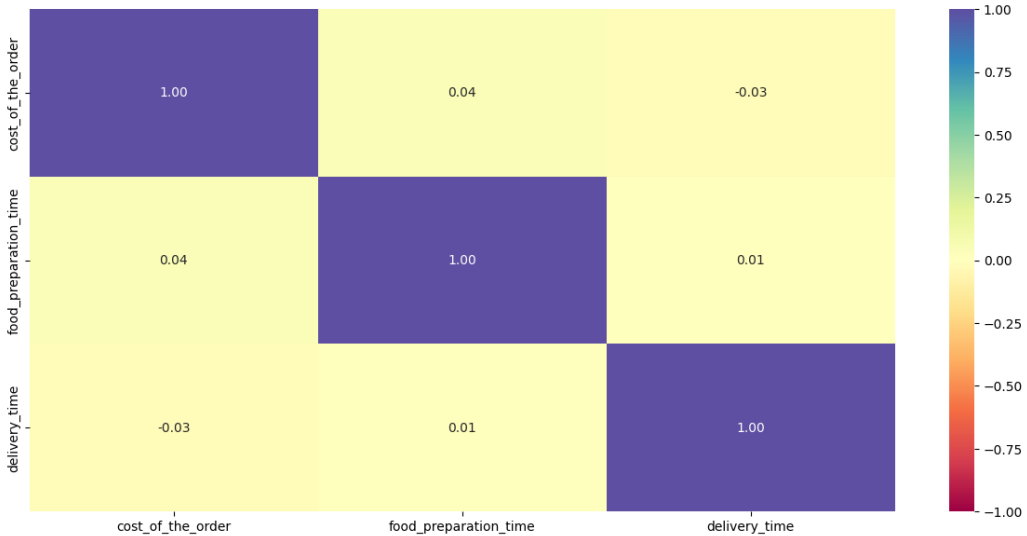
Multivariate Analysis

- **Question 12**: Rating vs. Cost of the Order
  - Rating 5: Highest average cost (~$17), suggesting customers who spend more tend to give top ratings.
  - Rating 3: Drops back to around $16.2, indicating no strong correlation between cost and lower ratings.
  - Rating 4: Slightly higher than “Not given” (~$16.8 vs. $16.1).



Multivariate Analysis

- **Question 12**: Correlation among variables
    - Cost of Order and Food Prep Time have a higher correlation (positive), but not significant
    - Higher order cost does not imply longer preparation or delivery time.
    - Preparation time and delivery time are almost independent.



Multivariate Analysis

- **Question 13**: The company wants to provide a promotional offer in the advertisement of the restaurants. The condition to get the offer is that the restaurants must have a rating count of more than 50 and the average rating should be greater than 4.
    - Find the restaurants fulfilling the criteria to get the promotional offer.
    - The table on the right shows the restaurants that meet the criteria.

	restaurant_name	rating
0	The Meatball Shop	4.511905
1	Blue Ribbon Fried Chicken	4.328125
2	Shake Shack	4.278195
3	Blue Ribbon Sushi	4.219178

Multivariate Analysis

- Question 14: The company charges the restaurant 25% on the orders having cost greater than 20 dollars and 15% on the orders having cost greater than 5 dollars. Find the net revenue generated by the company across all orders.
    - Net revenue is the Cost of the Order x by the corresponding %, and the sum.
    - The net revenue is around 6166.3 dollars

	order_id	customer_id	restaurant_name	cuisine_type	cost_of_the_order	day_of_the_week	rating	food_preparation_time	delivery_time	Revenue
0	1477147	337525	Hangawi	Korean	30.75	Weekend	Not given	25	20	7.6875
1	1477685	358141	Blue Ribbon Sushi Izakaya	Japanese	12.08	Weekend	Not given	25	23	1.8120
2	1477070	66393	Cafe Habana	Mexican	12.23	Weekday	5	23	28	1.8345
3	1477334	106968	Blue Ribbon Fried Chicken	American	29.20	Weekend	3	25	15	7.3000
4	1478249	76942	Dirty Bird to Go	American	11.59	Weekday	4	25	24	1.7385

Multivariate Analysis

- **Question 15:** The company wants to analyze the total time required to deliver the food. What percentage of orders take more than 60 minutes to get delivered from the time the order is placed? (The food has to be prepared and then delivered.)
    - **The percentage of orders with more than 60 minutes of total delivery time is 10.54%**

- **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?
    - **The mean delivery time on weekdays is around 28 minutes**
    - **The mean delivery time on the Weekend is around 22 minutes**

Multivariate Analysis

- **Question 17:** What are your conclusions from the analysis? What recommendations would you like to share to help improve the business? (You can use cuisine type and feedback ratings to drive your business recommendations.)
  - **Low Correlation Across Metrics**
    - Cost, preparation time, and delivery time are largely independent.
    - Rating has a weak positive correlation with cost & preparation time, and almost null with delivery time.
  - **Cuisine Insights**
    - **Mexican & American**: High ratings, moderate prep time.
    - **Italian & Indian**: Longer prep times, slightly lower ratings; need process optimization.
    - **Chinese**: Moderate ratings, fast prep; good for quick delivery campaigns.
  - **Customer Feedback:**
    - Ratings are not strongly tied to delivery speed, suggesting food quality and experience matter more than speed beyond a certain threshold.

Multivariate Analysis

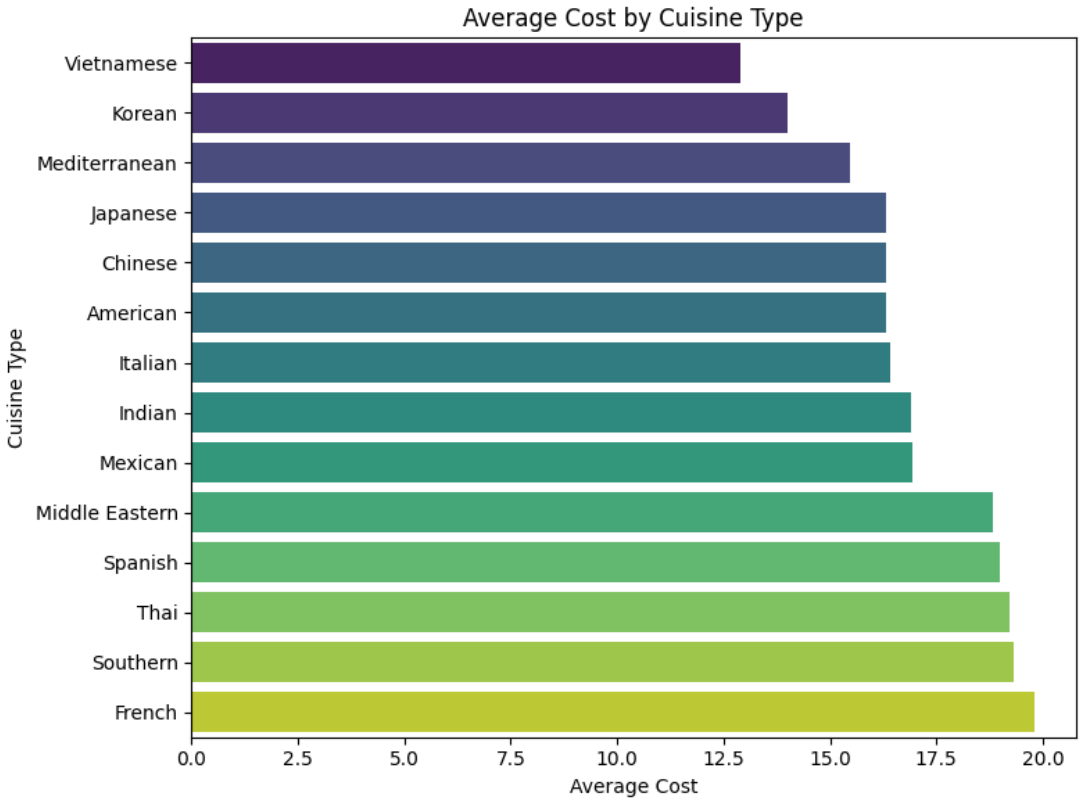
- Recommendations:
  - **Operational Efficiency**
    - Optimize delivery routes using predictive models (traffic, distance), particularly on week days.
    - Promote restaurants with lower preparation times to help increase productivity
  - **2. Menu & Cuisine Strategy**
    - Promote cuisines with high ratings and low prep time (Mexican, Spanish, Thai).
    - For cuisines with long prep times (Italian, Indian), consider: Pre-prep strategies for popular dishes & Limited-time menus focusing on faster dishes.
  - **3. Customer Experience**
    - Implement real-time order tracking and accurate ETA.
    - Highlight top-rated dishes in the app to boost satisfaction.
  - **4. Marketing & Promotions**
    - Offer discounts or loyalty points for less popular cuisines to balance demand.
    - Use feedback-driven campaigns: “Top-rated dishes this week.”
  - **5. Data-Driven Improvements**
    - Continuously monitor cuisine performance and feedback.
    - Run tests for menu changes and route efficiency by day/time for delivery optimizations.

APPENDIX

Slide Header

- To improve revenue, it is recommended to promote cuisine and restaurants with a higher cost of order

- French, Southern and Thai can bring higher revenue. Currently American cuisine brings higher revenue due to the overall volume, but changing the mix can increase net revenue for the company.



Great Learning

Happy Learning !