



FoodHub Data Analysis

Michail Mersinias

01/19/2023



Contents / Agenda

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

Executive Summary

- The executive summary (Q17) of the FoodHub Data Analysis presentation is as follows (1/2):
 - **Cuisines Types:** American, Japanese, Italian and Chinese are the most popular. Furthermore, they are the only cuisine types represented in the top 14 restaurants with the highest revenue.
 - **Cost of the Order:** The majority (75%) of the orders cost between \$12 and \$22.
 - **Daily Seasonality:** The daily order rate is 6.17 times higher during the weekends.
 - **Unrated Restaurants:** An issue is that 38.77% of the restaurants remain unrated.
 - **The Big Few:** The top 5 restaurants correspond to 33.4% of the total orders. Furthermore, they are the only one who achieved a revenue greater than \$1000.
 - **Cheap and Fast:** Korean and Vietnamese cuisines display statistically significant lower order cost and food preparation time, when compared to the rest.
 - **Rating Matters:** Rating is less likely to be low when the delivery time is low and the order cost is high. It is more likely for a customer to provide a rating when the order cost is high.

Executive Summary

- The executive summary (Q17) of the FoodHub Data Analysis presentation is as follows (2/2):
 - **Recommendation 1:** Customers seem to strongly prefer American, Japanese, Italian and Chinese food. FoodHub should attempt to attract more such restaurants, by targeting them with a promotional campaign which, for a short time, offers them a discounted 10% fee.
 - **Recommendation 2:** Customers seem to strongly prefer the weekend over the weekdays. However, since the demand exists, FoodHub should attempt to provide a promotional campaign targeting customers during the weekdays, by offering to cover the delivery fee cost.
 - **Recommendation 3:** Customers seem partially unwilling to provide ratings, especially on orders with a low cost. As a result, FoodHub should not only take rating into account to provide restaurants with a promotional offer, but also other factors such as delivery time.
 - **Recommendation 4:** The revenue generation policy of FoodHub is problematic as very few orders cost over \$25. A policy of a static 20% fee on all orders would provide more revenue for FoodHub, and would incentivize the restaurants to aim for orders over \$25 as well.

Business Problem Overview and Solution Approach

- Problem Definition:
 - FoodHub, a food aggregator company, provides the data of different orders made by the customers to the restaurants, in their online portal.
 - The goal is to find insights from data analysis in order to provide recommendations to FoodHub, which will be aimed to increase its revenue generation as well as both restaurant and customer retention and growth.
- Solution Approach and Methodology:
 - Initial exploratory data analysis and overview will be performed to see the big picture.
 - For each attribute, univariate analysis will be performed to understand its distribution.
 - Multivariate analysis will be performed to examine the relationships between attributes.
 - Recommendations will be given, based on both quantitative and qualitative observations as well as insights from the aforementioned data analysis.

Data Overview

- In this section, we will proceed perform the data overview and initial exploratory data analysis.
 - We will report the shape of the dataset
 - We will examine each column data description and data type
 - We will provide a report on potential missing or problematic values
 - We will provide some quick observations and insight from the data.

Data Overview

- The dataset is comprised of 1898 rows and 9 columns. The columns are as follows:

•	order_id	Unique id of Order	Type: Int64
	customer_id	Customer id	Type: Int64
	restaurant_name	Name of Restaurant	Type: Object (String/Text)
	cuisine_type	Type of Cuisine	Type: Object (String/Text)
	cost_of_the_order	Cost of the Order (Dollars)	Type: Float64
	day_of_the_week	Weekday or Weekend	Type: Object (String/Text)
	rating	Rating (scale of 5 stars)	Type: Object (String/Text)
	food_preparation_time	Preparation time (minutes)	Type: Int64
	delivery_time	Delivery Time (minutes)	Type: Int64

Data Overview

- The dataset has no missing values in any of the columns, thus no action to replace missing values is necessary. We do not, however, that the “Rating” column has values out of the designated scale of 5 stars, which are labelled as “Not Given”.
- Some quick observations and insights from the data:
 - Food Preparation Time
 - Min = 20
 - Mean = 27.37
 - Max = 35
 - Unrated Orders (Rating = “Not Given”)
 - Total = 736

Univariate Analysis

- In this section, we perform univariate analysis on the data.
 - For each attribute, we perform a descriptive statistical analysis where only that attribute is involved as a variable.
 - We also analyze the corresponding data and present the distribution of the attribute, with confidence intervals to signify variance and statistical significance.
 - Finally, we write the conclusion based on both quantitative and qualitative observations.
- Furthermore, Q6 to Q11 are addressed in this section.

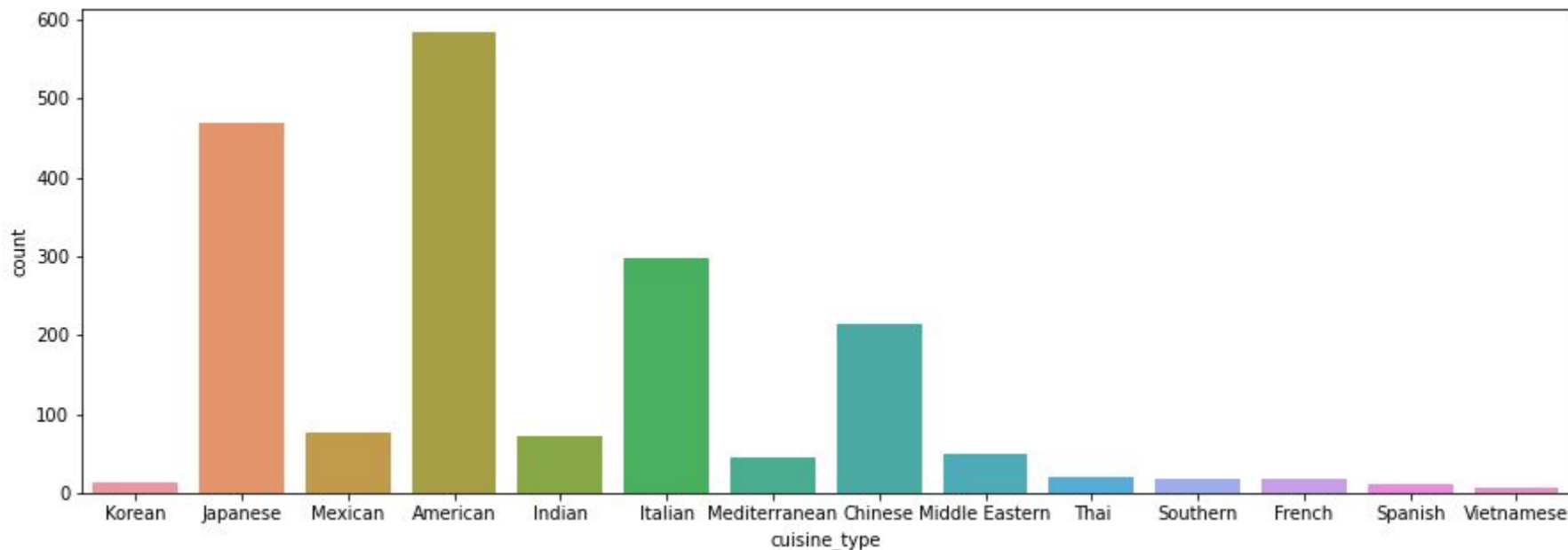
Univariate Analysis

- Unique Values:

Dataset Column	Unique Values
order_id	1898
customer_id	1200
restaurant_name	178
cuisine_type	14
cost_of_the_order	312
day_of_the_week	2
rating	4
food_preparation_time	16
delivery_time	19

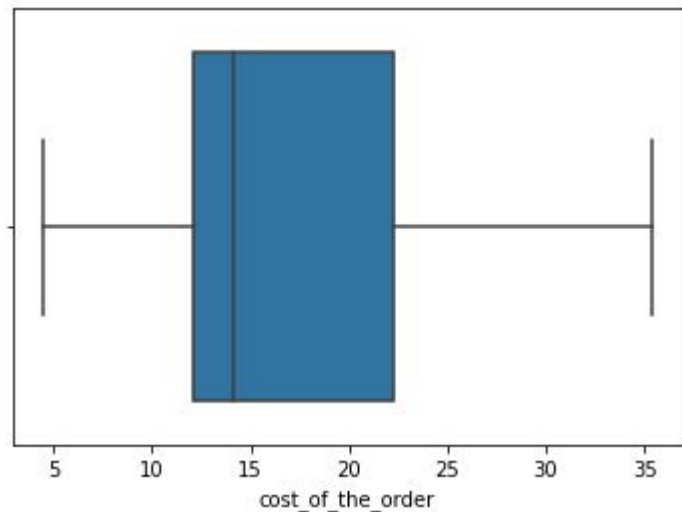
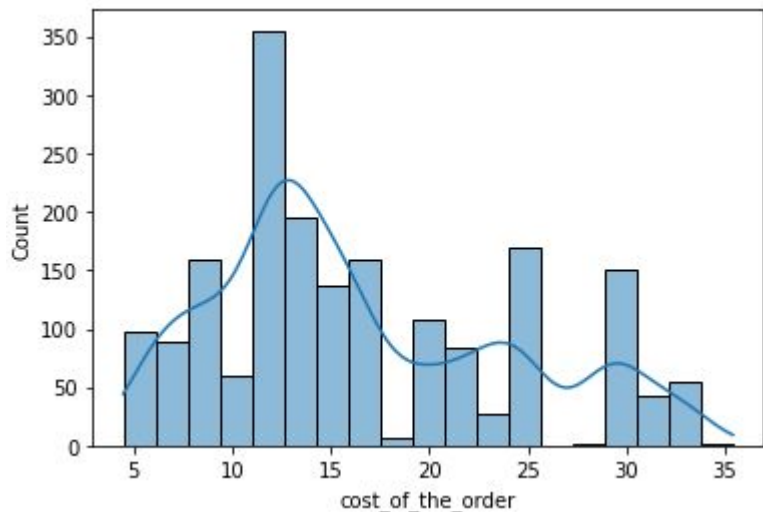
Univariate Analysis

- Analyzing the `cuisine_type` attribute, we observe that American cuisine is the most popular with 584 orders. It is followed by the Japanese (470 orders), Italian (298 orders) and Chinese (215 orders) cuisines. The rest have a count of less than 100 orders. A countplot that depicts this follows.



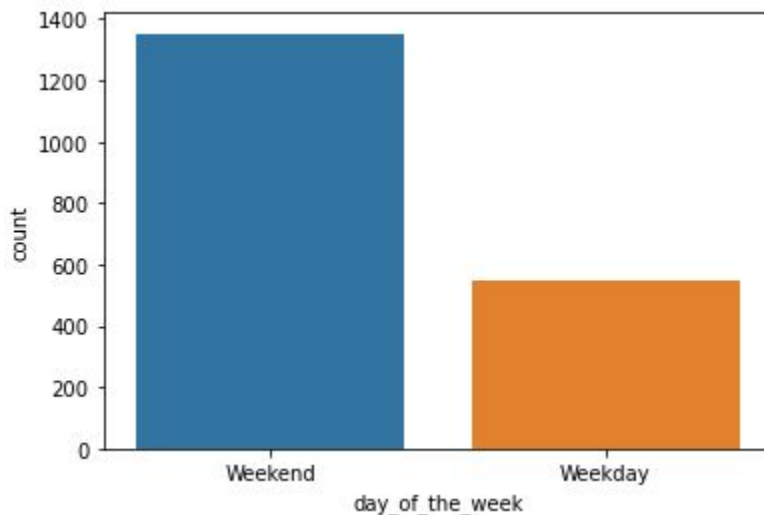
Univariate Analysis

- Analyzing the `cost_of_the_order` attribute, we report the min cost (\$4.47), the median cost (\$14.14), the mean cost (\$16.50) and the max cost (\$35.41), with a standard deviation of 7.48. Two graphs, a histplot and a boxplot, are presented to depict this more clearly.



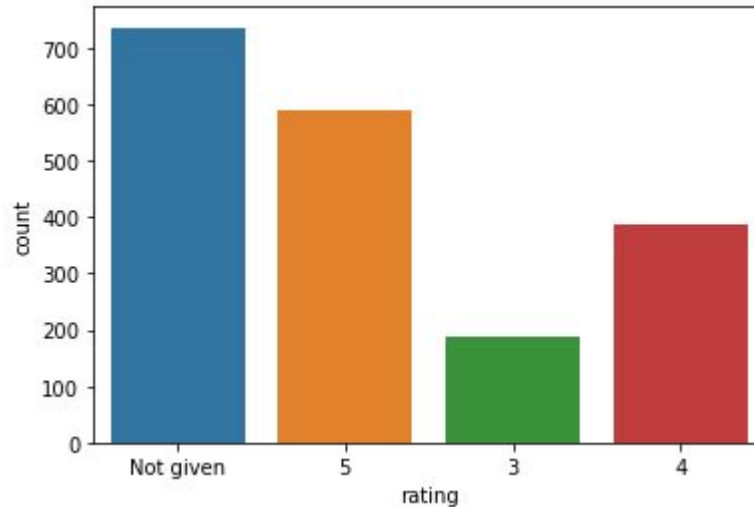
Univariate Analysis

- Analyzing the `day_of_the_week` attribute, we observe 1351 orders during the weekend and 547 orders during a weekday. Considering that a weekend only has 2 days, that brings the daily weekend order rate to 675.5 orders per day. On the other hand, for the other 5 weekdays, the daily weekday order rate is 109.4. Thus, we emphasize that the weekend displays high daily seasonality!



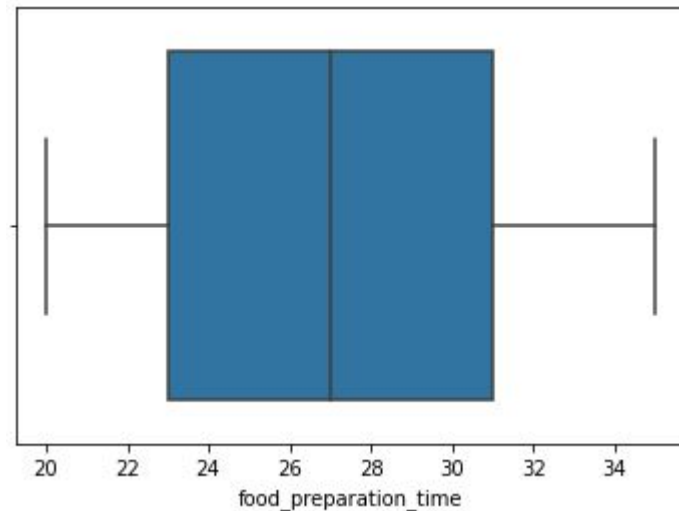
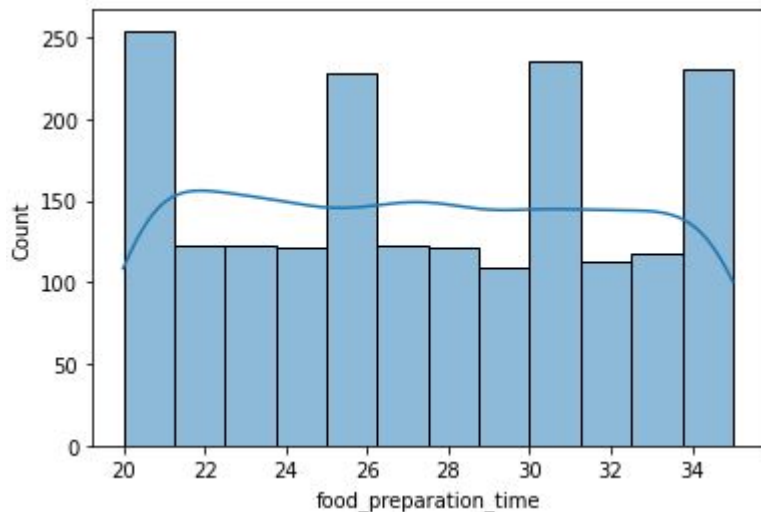
Univariate Analysis

- Analyzing the rating attribute, we observe that a significant number of the orders (736) were not given a rating. The rest (1162) which were assigned a rating, displayed high rating. The maximum rating of 5 was assigned 588 times, a rating of 4 stars was assigned 386 times and a rating of 3 stars was assigned 188 times. The average rating was 4.34 stars.



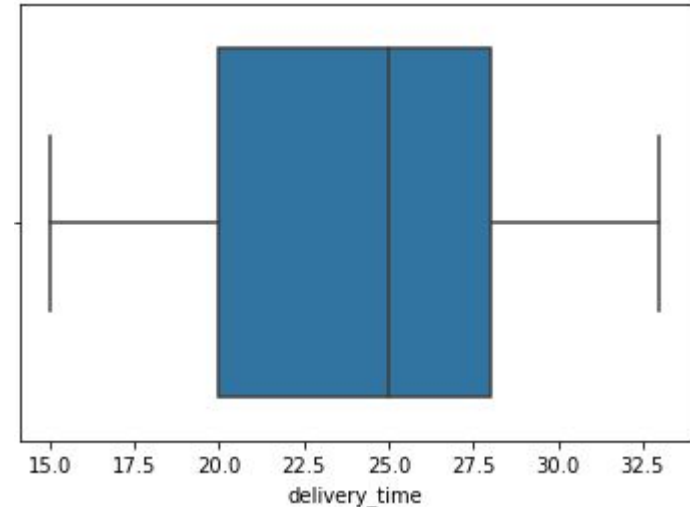
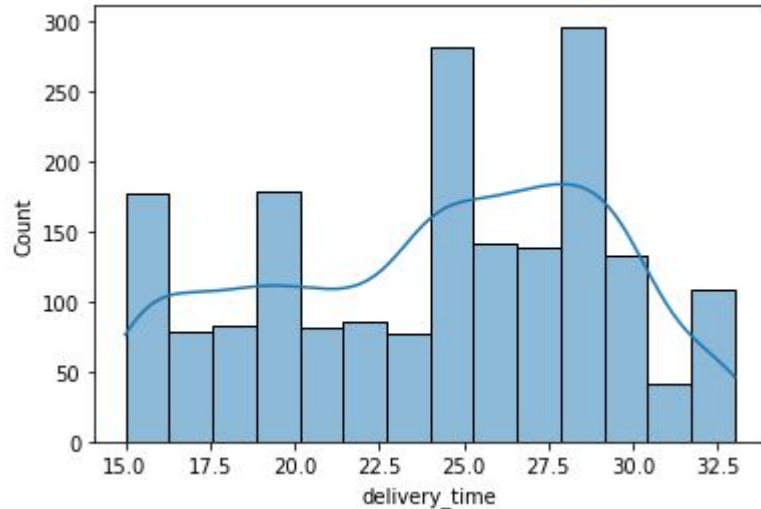
Univariate Analysis

- Analyzing the `food_preparation_time` attribute, we report the min food preparation time (20 minutes), the median (27 minutes), the mean (27.37 minutes) and the max (35 minutes), with a standard deviation of 4.63. Two graphs, a histplot and a boxplot, are presented to depict this.



Univariate Analysis

- Analyzing the `delivery_time` attribute, we report the min food preparation time (15 minutes), the median (25 minutes), the mean (24.16 minutes) and the max (33 minutes), with a standard deviation of 4.97. Two graphs, a histplot and a boxplot, are presented to depict this more clearly.



Univariate Analysis

- A few more insights from our univariate analysis:
 - Top 5 restaurant names with the largest number of orders:
 - Shake Shack (219 orders)
 - The Meatball Shop (132 orders)
 - Blue Ribbon Sushi (119 orders)
 - Blue Ribbon Fried Chicken (96 orders)
 - Parm (68 orders)
 - Most popular cuisine on the weekends:
 - American (415 orders on the weekends)

Univariate Analysis

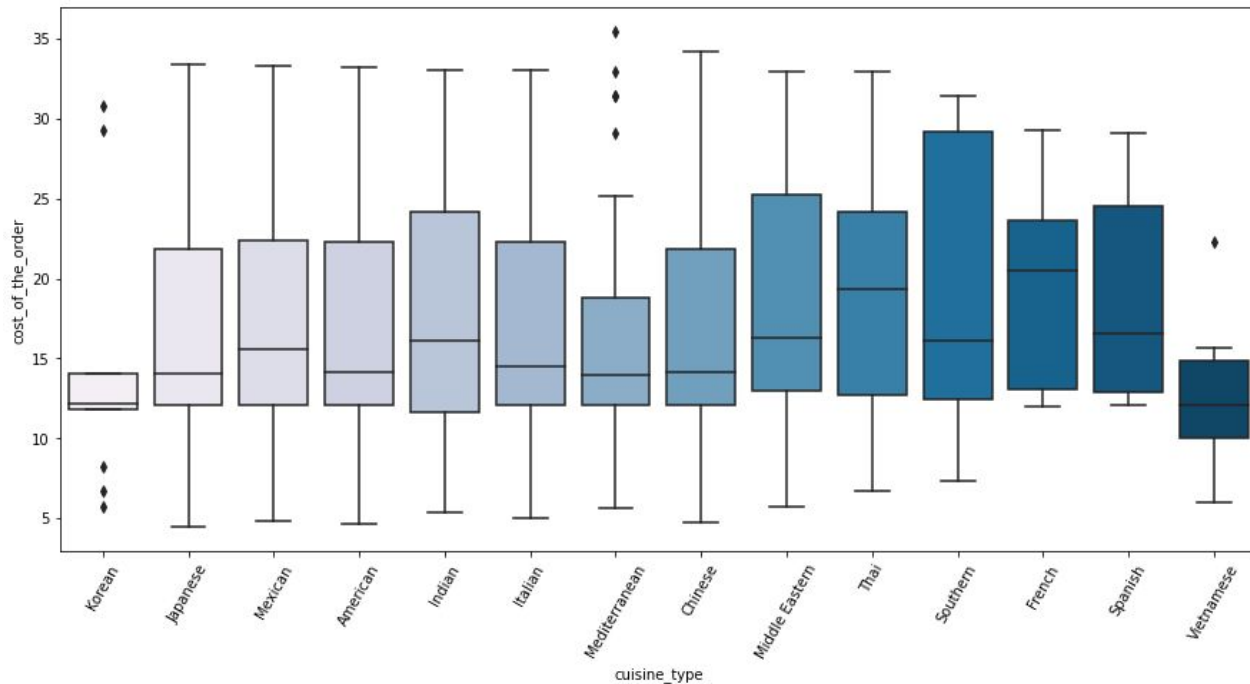
- A few extra insights from our univariate analysis:
 - The percentage of orders with a cost more than \$20 is 29.24%.
 - Delivery Order Time:
 - Mean: 24.16 minutes
 - Top 3 most frequent customers, eligible for 20% discount vouchers:
 - Customer ID 52832 , with 13 orders placed
 - Customer ID 47440 , with 10 orders placed
 - Customer ID 83287 , with 9 orders placed

Multivariate Analysis

- In this section, we perform multivariate analysis on the data.
 - For selected attributes, we perform a descriptive statistical analysis where each attribute is jointly examined along with other attributes in order to evaluate their relationship and correlation.
 - We also analyze the corresponding data and present the joint distribution of the attributes, with confidence intervals to signify variance and statistical significance.
 - Finally, we write the conclusion based on both quantitative and qualitative observations.
- Furthermore, Q12 to Q16 are addressed in this section.

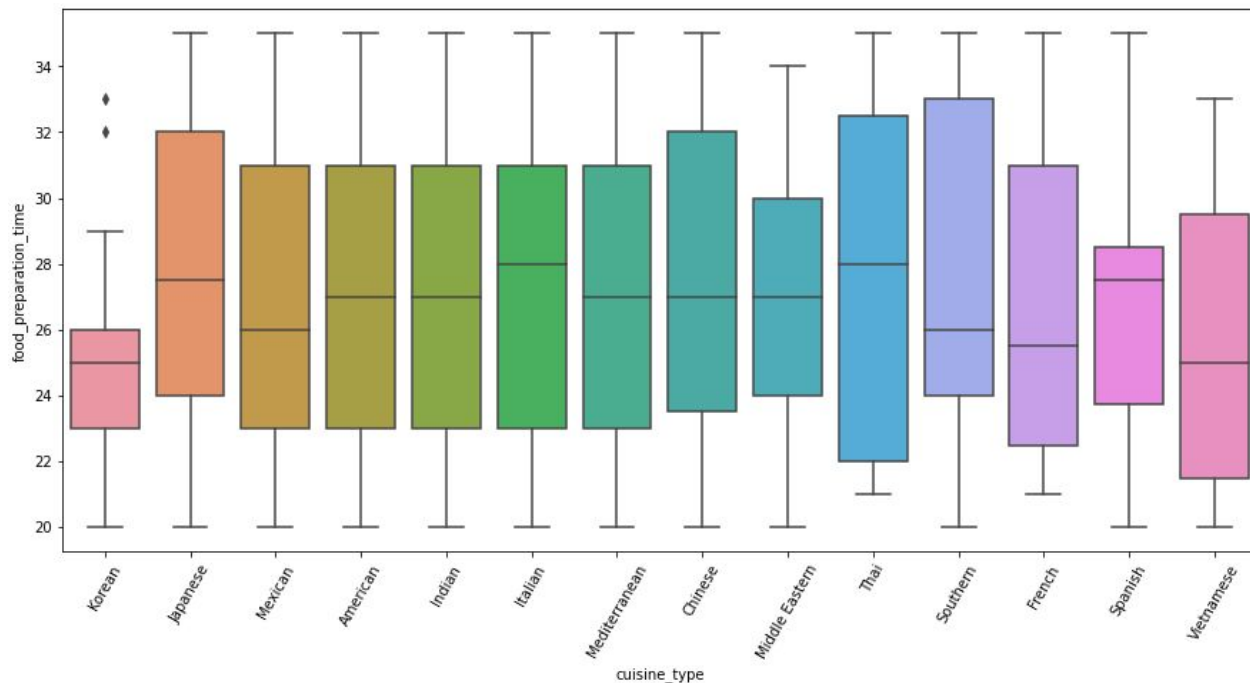
Multivariate Analysis

- Analyzing the relationship between the `cuisine_type` and `cost_of_the_order` attributes, we observe that the Korean, Mediterranean and Vietnamese cuisines display a statistically significant lower cost than the rest, due to their lower median values and low variance, despite their few outliers.



Multivariate Analysis

- Analyzing the relationship between the `cuisine_type` and `food_preparation_time` attributes, we observe that the Korean and Vietnamese cuisines display a statistically significant lower preparation time than the rest, due to their lower median values and (relatively to others) low variance.



Multivariate Analysis

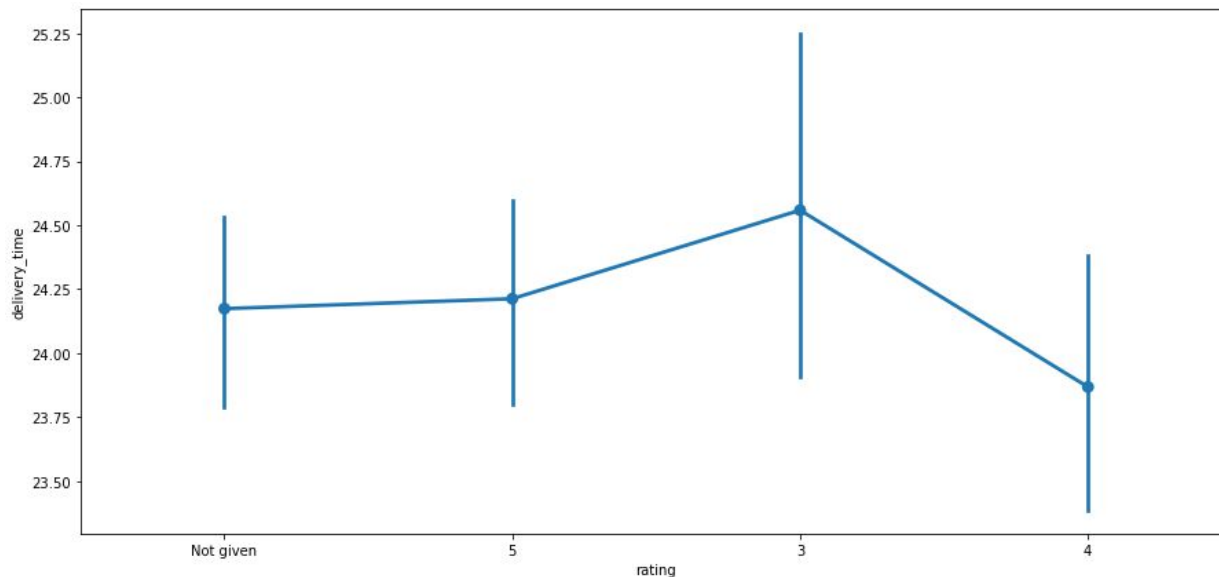
- Analyzing the relationship between the restaurants which generated the largest revenue (Top 1-7) and their respective cuisine_type attribute, we present our results:
 1. Shake Shack (American, \$3579, 53)
 2. The Meatball Shop (85% Italian - 15% American, \$2145.21)
 3. Blue Ribbon Sushi (Japanese, \$1903.95)
 4. Blue Ribbon Fried Chicken (American, \$1662.29)
 5. Parm (Italian, \$1112.76)
 6. RedFarm Broadway (Chinese, \$965.13)
 7. RedFarm Hudson (Chinese, \$921.21)

Multivariate Analysis

- Analyzing the relationship between the restaurants which generated the largest revenue (Top 8-14) and their respective cuisine_type attribute, we present our results:
 8. TAO (Japanese, \$834.50)
 9. Han Dynasty (Chinese, \$755.29)
 10. Blue Ribbon Sushi Bar & Grill (Japanese, \$666.62)
 11. Rubirosa (Italian, \$660.45)
 12. Sushi of Gari 46 (Japanese, 640.87)
 13. Nobu Next Door (Japanese, 623.67)
 14. Five Guys Burgers and Fries (American, 506.47)

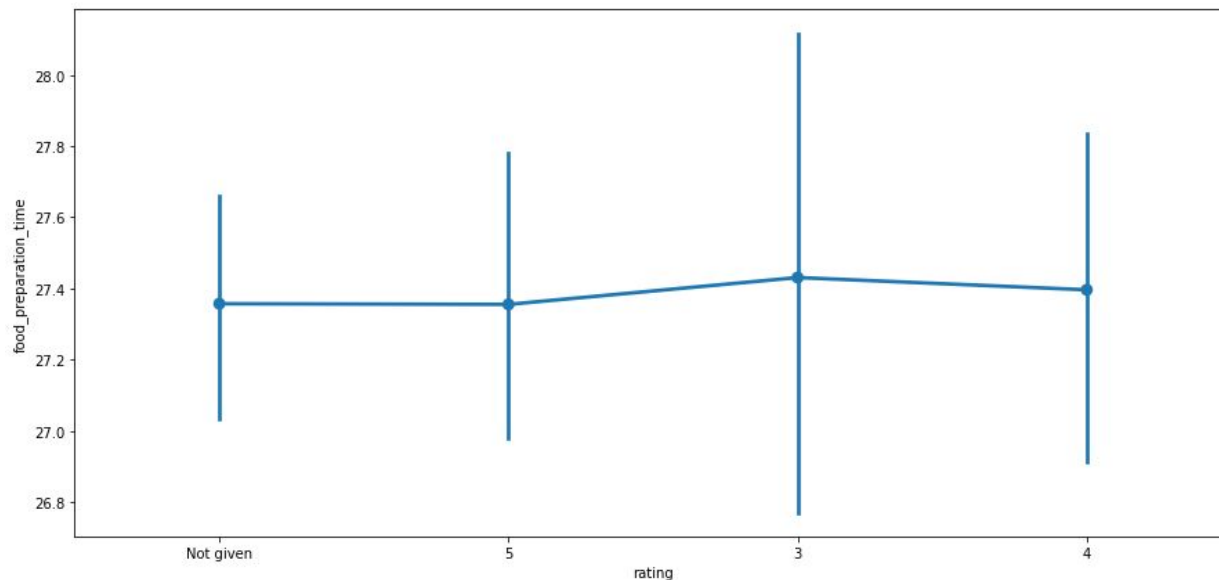
Multivariate Analysis

- Analyzing the relationship between the rating and delivery_time attributes, we observe that lower delivery times display less variance and are more likely to avoid a bad (3 stars) rating.



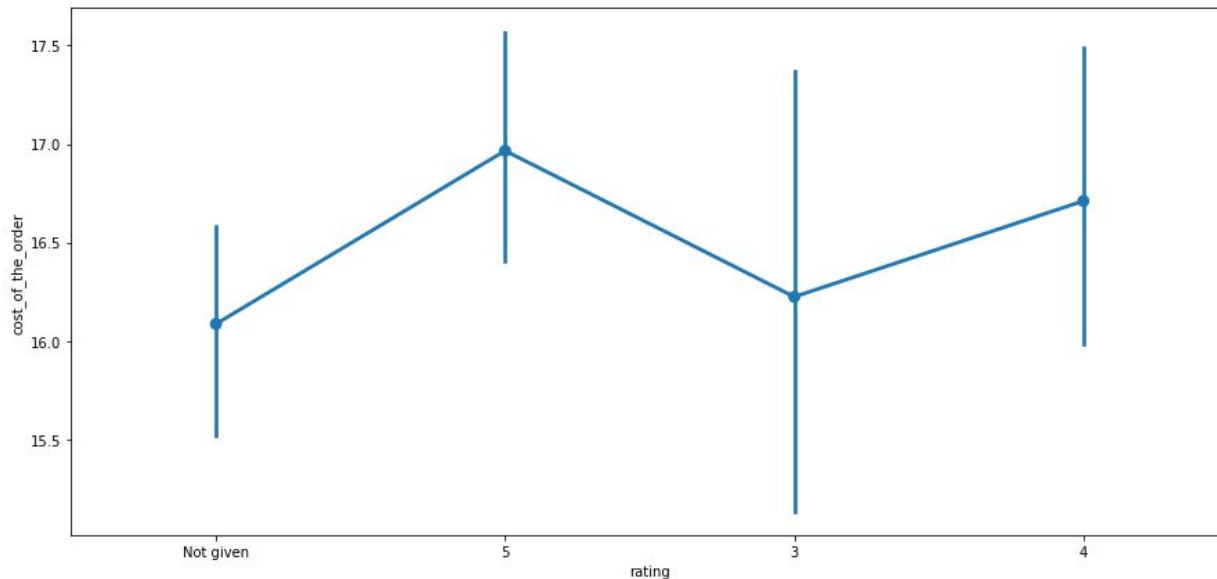
Multivariate Analysis

- Analyzing the relationship between the rating and food_preparation_time attributes, we observe that lower food preparation times display less variance, but otherwise do not significantly affect the given rating.



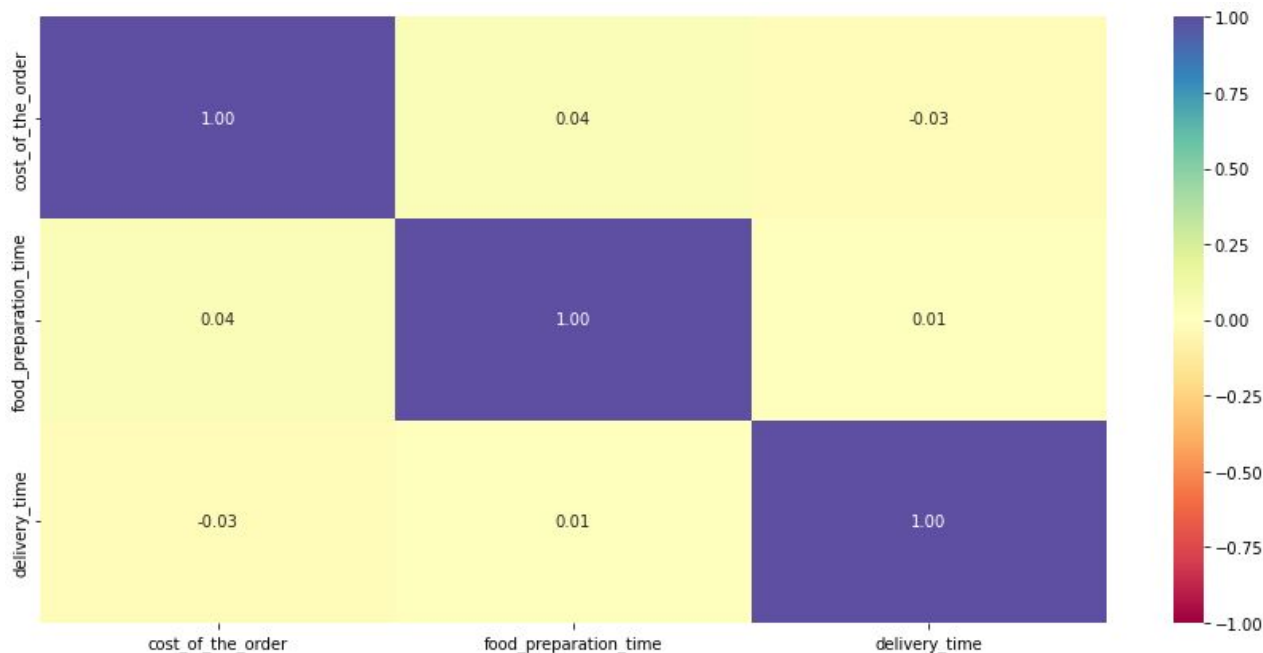
Multivariate Analysis

- Analyzing the relationship between the rating and cost_of_the_order attributes, we observe that a higher order cost leads to less variance and is more likely to avoid both the absence of a rating and a bad (3 stars) rating.



Multivariate Analysis

- Examining the correlation between the three attributes (delivery_time, food_preparation_time and cost_of_the_order), we observe that there is no particular strong correlation between them.



Multivariate Analysis

- The promotional offer of the company is conditional on the restaurants having more than 50 ratings and an average rating score greater than 4 stars. Thus, the following restaurants qualify:
 1. The Meatball Shop, with 84 ratings and an average rating score of 4.51
 2. Blue Ribbon Fried Chicken, with 64 ratings and an average rating score of 4.33
 3. Shake Shack, with 133 ratings and an average rating score of 4.28
 4. Blue Ribbon Sushi, with 73 ratings and an average rating score of 4.22

Multivariate Analysis

- A few more insights from our multivariate analysis:
 - The policy which charges the restaurants 25% on the orders having cost greater than \$20 and 15% on the orders having cost greater than \$5, generated a net revenue of \$6166.3 for the company in total.
 - The total time to deliver food, defined as the sum of food preparation time and delivery time, is more than 60 minutes for 12.89% of the total orders.
 - The mean delivery time varies between weekends and weekdays. More specifically:
 - Mean delivery time on weekdays: 28 minutes.
 - Mean delivery time on weekends: 22 minutes.

Executive Summary

- The executive summary (Q17) of the FoodHub Data Analysis presentation is as follows (1/2):
 - **Cuisines Types:** American, Japanese, Italian and Chinese are the most popular. Furthermore, they are the only cuisine types represented in the top 14 restaurants with the highest revenue.
 - **Cost of the Order:** The majority (75%) of the orders cost between \$12 and \$22.
 - **Daily Seasonality:** The daily order rate is 6.17 times higher during the weekends.
 - **Unrated Restaurants:** An issue is that 38.77% of the restaurants remain unrated.
 - **The Big Few:** The top 5 restaurants correspond to 33.4% of the total orders. Furthermore, they are the only one who achieved a revenue greater than \$1000.
 - **Cheap and Fast:** Korean and Vietnamese cuisines display statistically significant lower order cost and food preparation time, when compared to the rest.
 - **Rating Matters:** Rating is less likely to be low when the delivery time is low and the order cost is high. It is more likely for a customer to provide a rating when the order cost is high.

Executive Summary

- The executive summary (Q17) of the FoodHub Data Analysis presentation is as follows (2/2):
 - **Recommendation 1:** Customers seem to strongly prefer American, Japanese, Italian and Chinese food. FoodHub should attempt to attract more such restaurants, by targeting them with a promotional campaign which, for a short time, offers them a discounted 10% fee.
 - **Recommendation 2:** Customers seem to strongly prefer the weekend over the weekdays. However, since the demand exists, FoodHub should attempt to provide a promotional campaign targeting customers during the weekdays, by offering to cover the delivery fee cost.
 - **Recommendation 3:** Customers seem partially unwilling to provide ratings, especially on orders with a low cost. As a result, FoodHub should not only take rating into account to provide restaurants with a promotional offer, but also other factors such as delivery time.
 - **Recommendation 4:** The revenue generation policy of FoodHub is problematic as very few orders cost over \$25. A policy of a static 20% fee on all orders would provide more revenue for FoodHub, and would incentivize the restaurants to aim for orders over \$25 as well.



Thank you for your time!

Michail Mersinias

01/19/2023