

# FOODHUB DELIVERY SERVICE ANALYSIS REPORT

MIT Applied Data Science Program

Project assessment: Foundations for Data Science

Delia McNamara

3 June 2024

## Contents / Agenda

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

## Executive Summary

- Analyze Foodhub Delivery Service restaurant demand for variables relating to customer experience.
- Net revenue was \$6,166.30.
  - ◇ Shake Shack highest revenue restaurant at \$3,579.53.
- 4 restaurants rated highly and frequently enough for proposed promotion.
  - ◇ The Meatball Shop top performer by rating frequency & value.
- American cuisine most popular cuisine type on weekends.
- Higher order cost associated with higher rating.
- 40% of orders unrated and low correlation between rating value and the variables used in these analyses.
  - ◇ Recommend further analysis with other customer satisfaction metrics for better insight.
  - ◇ Alternatively, target increased frequency of customer ratings for better results.
- Weekend orders represent a significant portion of the total orders.
  - ◇ Targeting this sector of the data with new strategies will easily target a large proportion of orders.

# Business Problem Overview and Solution Approach

→ Food delivery app Foodhub is seeking to analyze order data to gain insight on the performance of different restaurants on the platform.

- ◇ Goal is to enhance customer experience.

→ Visualize data & analyses to develop insight into individual variables:

- ◇ Order ID: unique ID numbers assigned to each order.
- ◇ Customer ID: unique ID numbers assigned to each customer.
- ◇ Restaurant name: name of the restaurant.
- ◇ Cuisine type: cuisine type of the restaurant.
- ◇ Order cost: total cost of the order placed by the customer.
- ◇ Day of the week: weekday (Monday to Friday) or weekend (Saturday & Sunday).
- ◇ Rating: rating score given by the customer, out of 5 points.
- ◇ Food prep time: number of minutes taken by the restaurant to prepare the food.
- ◇ Delivery time: number of minutes taken by the delivery person to deliver the order to the customer from the restaurant.

→ Perform univariate analysis to determine:

- ◇ Quantity of unique values within variables.
- ◇ Order frequencies by individual variables.
- ◇ Cuisine type preferences on certain days of the week.
- ◇ Highest performing restaurants by order frequency.
- ◇ Customers eligible for 20% discount vouchers.

→ Perform multivariate analysis to determine:

- ◇ Relationships between key variables for customer experience.
- ◇ Highest revenue restaurants & net revenue.
- ◇ Restaurants eligible for promotional offer.
- ◇ Insights into delayed orders & order frequency by day of the week.

## Data Overview

- ▲ **Data frame consisting of 1,898 rows & 9 columns.**

Variable	No. rows	Data type
Order ID	1,898	Integer
Customer ID	1,898	Integer
Restaurant name	1,898	Object
Cuisine type	1,898	Object
Order cost	1,898	Float
Day of the week	1,898	Object
Rating	1,898	Object
Food prep time	1,898	Integer
Delivery time	1,898	Integer

- ▲ Data from customer orders & feedback.
- ▲ All variables have 1,898 rows.
- ▲ Mix of data types:
  - ◇ Integer: 4 variables.
  - ◇ Object: 4 variables.
  - ◇ Float: 1 variable.

## Missing value treatment

- ▲ **No null or missing values for any variables.**
  - ◇ No missing value treatment needed.
- ▲ **Ratings not given:**
  - ◇ **736 orders ratings not given out of 1,898 orders.**
  - ◇ No treatment: customers chose not to rate.

## Summary statistics

Summary statistic	Cost of the order	Food prep time	Delivery time
Count	1898	1898	1898
Mean	\$16.50	27.4	24.2
Standard deviation	\$7.48	4.6	5.0
Minimum	\$4.47	20.0	15.0
25% quartile	\$12.08	23.0	20.0
50% quartile	\$14.14	27.0	25.0
75% quartile	\$22.30	31.0	28.0
Maximum	\$35.41	35.0	33.0

Summary statistics describing numeric variables in the dataset.

NOTE: Customer ID & Order ID are numeric variables, but are superfluous to the summary statistics, so were removed.

- ⤴ **Minimum food preparation time: 20.0 minutes.**
- ⤴ **Average food preparation time: 27.4 minutes.**
- ⤴ **Maximum food preparation time: 35.0 minutes.**
- ⤴ **Mean order delivery time: 24.2 minutes.**

# Univariate Analysis

→ Performed for all variables, categorical and numeric.

→ Patterns in distributions, visualization of variables.

## Order ID

- ▲ 1,898 unique order IDs.
  - ◇ One order per row.
  - ◇ 736 orders not rated.

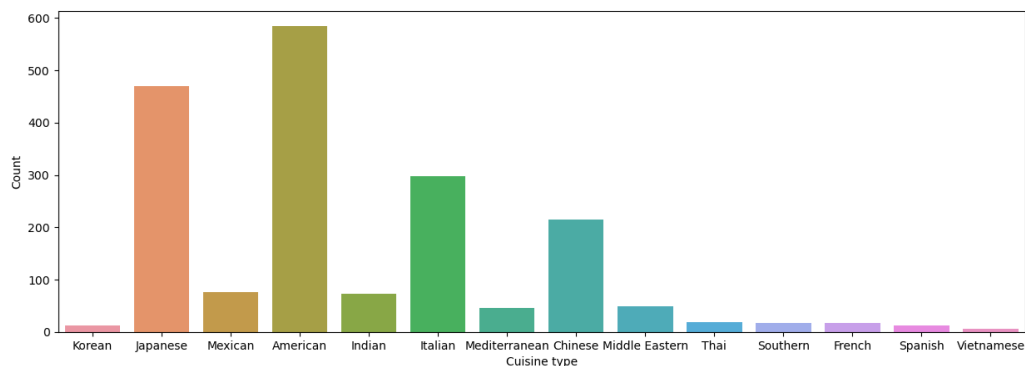
## Customer ID

- ▲ 1,200 unique customer IDs.
  - ◇ Lower than unique order IDs.
  - ◇ Indicates repeat customers.
  - ◇ Multiple orders per customer.

## Restaurant name

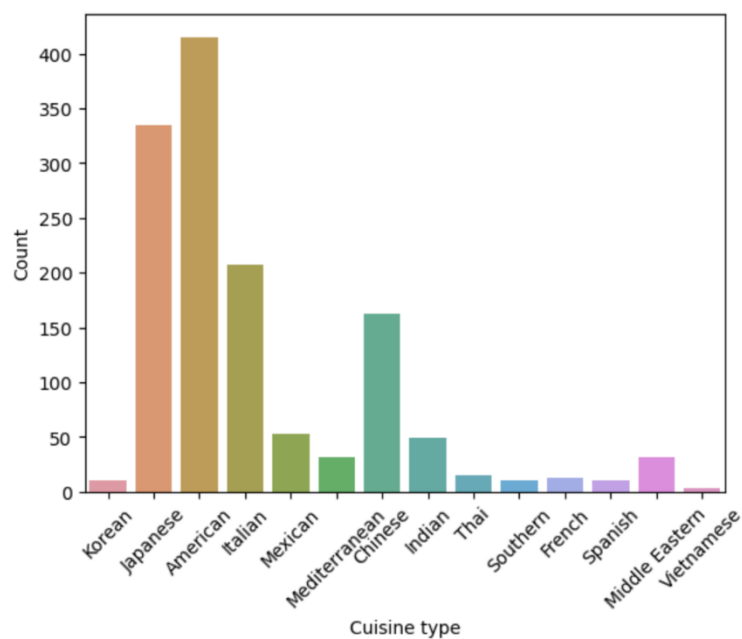
- ▲ 178 unique restaurant names.
  - ◇ Lower than both unique order & customer IDs.
  - ◇ Multiple orders per restaurant.
  - ◇ Potential repeat customers per restaurant.

## Cuisine preference



- ⤴ American cuisine most frequently ordered.
- ◊ Followed by Japanese, Italian, Chinese.
- ⤴ Vietnamese, Korean, Spanish least frequently ordered.

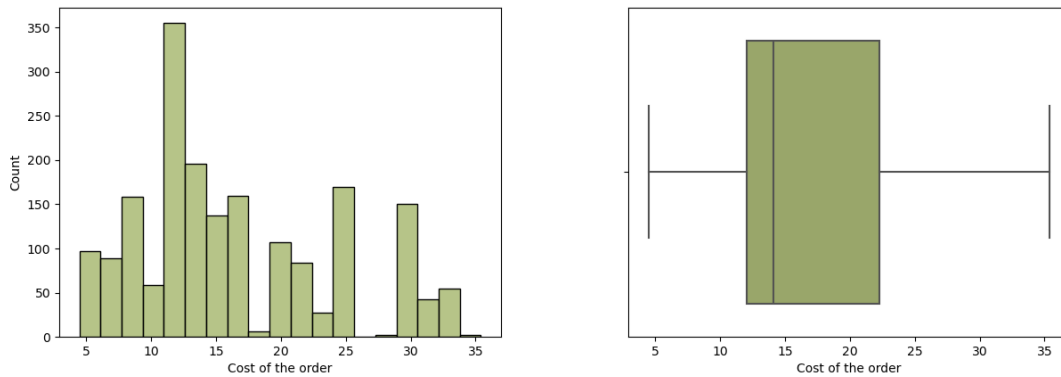
## Weekend cuisine preference



- ⤴ 14 unique types of cuisine ordered on the weekend.
- ⤴ Top 5 types of cuisine by number of orders:
  - ◊ American 415
  - ◊ Japanese 335
  - ◊ Italian 207
  - ◊ Chinese 163
  - ◊ Mexican 53

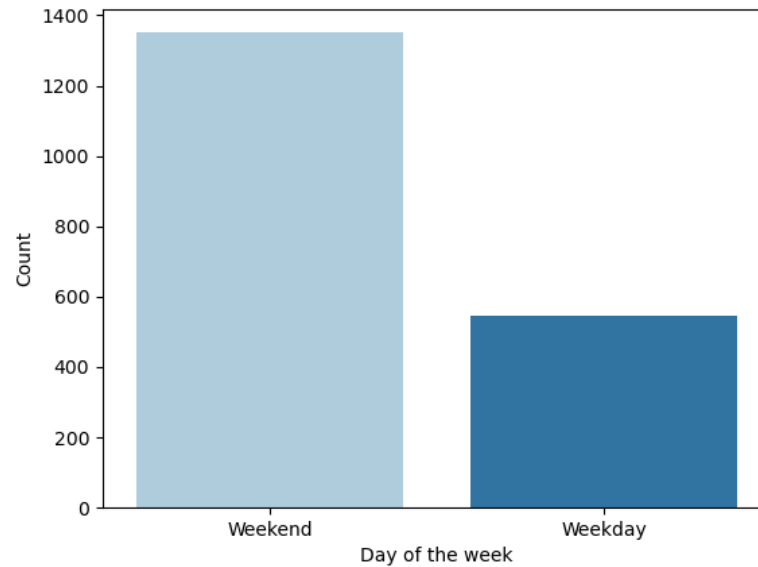


## Cost of the order



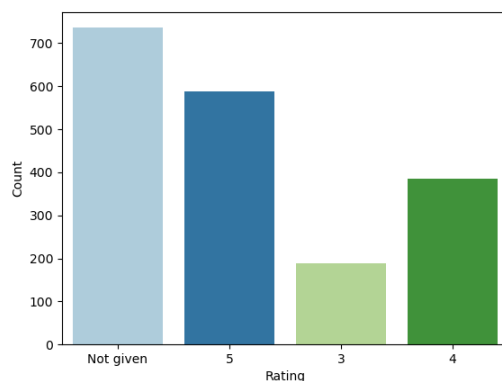
- ⬆ **Positive skew to order cost distribution.**
- ⬆ **No orders below \$5 or above \$35.**
- ⬆ **50% of the orders cost between \$12 to \$23.**
- ⬆ **Gap in distribution between \$25 to \$30.**
  - ◇ **Few order costs in this range.**
- ⬆ **No outliers in boxplot.**
- ⬆ **The number of orders that cost more than \$20 is 555.**
- ⬆ **The percentage of orders above \$20 is 29.24%.**

## Day of the week



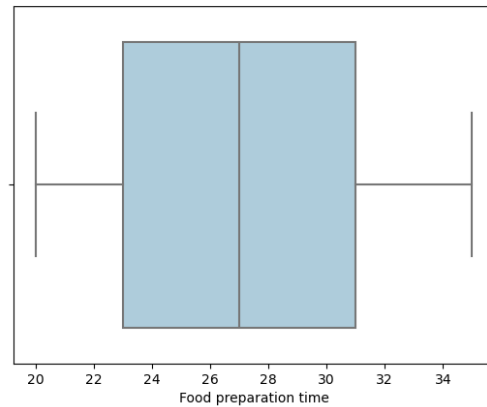
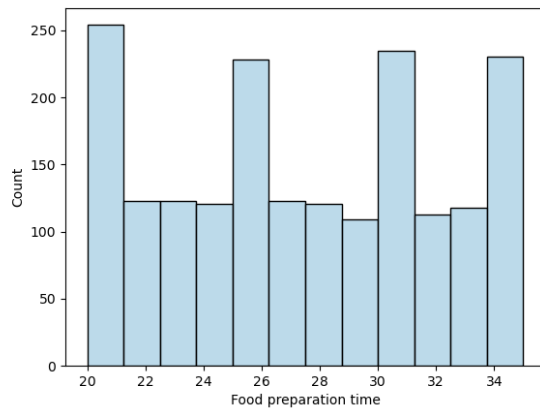
- ⤴ Over twice as many orders on weekends in comparison to weekdays.
- ⤴ Roughly 1,300 orders on weekends.
- ⤴ Roughly 500 orders on weekdays.

## Rating



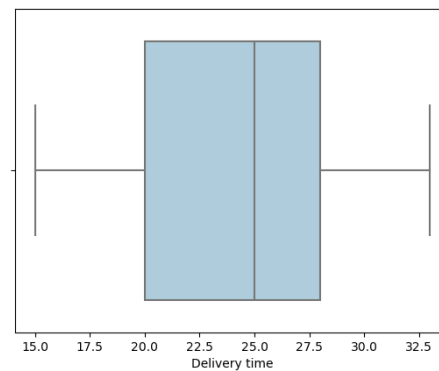
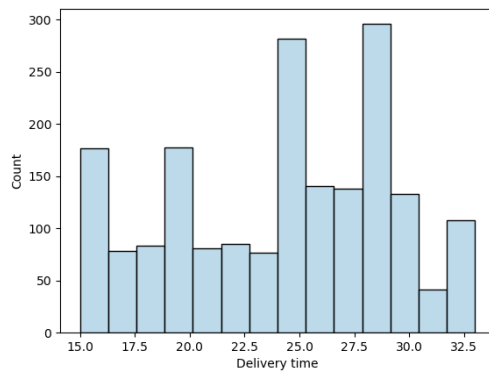
- ⤴ 4 unique values:
  - ◇ Not given, 5, 4, 3.
- ⤴ No ratings below 3.
- ⤴ Not given highest proportion.
- ⤴ 5 most frequent rating.
- ⤴ 3 least frequent rating.

## Food preparation time



- ⬆ **Uniform distribution with some high frequencies.**
- ⬆ **All values between 20 to 35 minutes.**
- ⬆ **No skewing nor outliers.**

## Delivery time



- ⬆ **Slight negative skew to delivery time distribution.**
- ⬆ **All values between 15 to 33 minutes.**
- ⬆ **No outliers in delivery time.**

## Top 5 restaurants

Restaurant	No. orders
Shake Shack	219
The Meatball Shop	132
Blue Ribbon Sushi	119
Blue Ribbon Fried Chicken	96
Parm	68

- ⤴ By total number of orders.
- ⤴ Shake shack highest number of orders.
  - ◇ Roughly 80 orders above next-highest.
- ⤴ Lesser magnitude difference between other entries.
- ⤴ Range of over 150 orders in top 5.

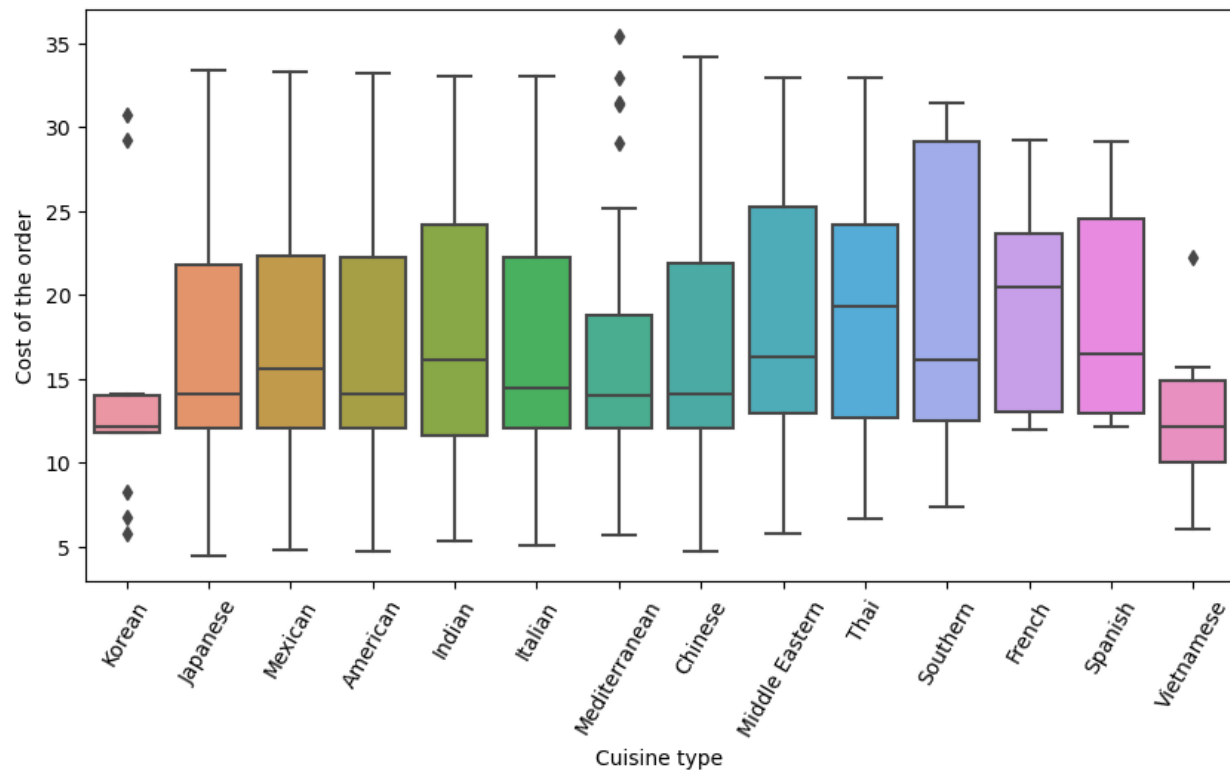
## 20% discount vouchers

Top 3 customers	No. orders
52832	13
47440	10
83287	9

- ⤴ Top 3 customers with the most orders will be given a 20% discount voucher.

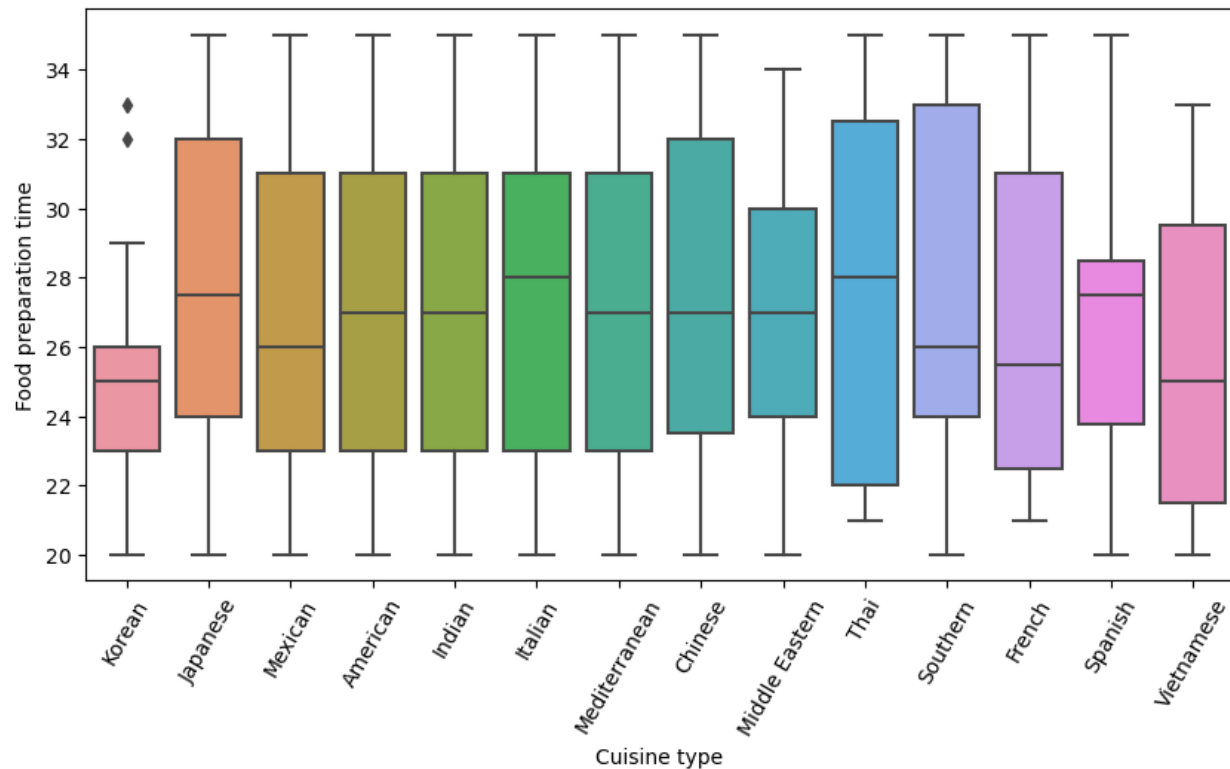
# Multivariate Analysis

## Cuisine type & order cost



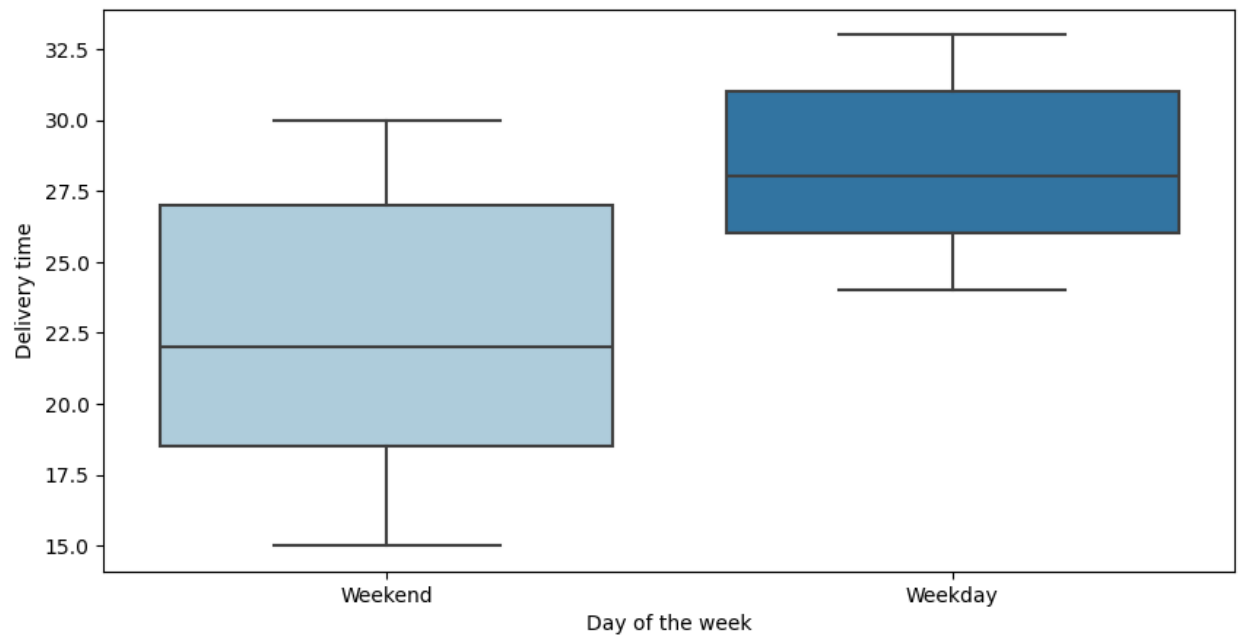
- ⤴ Most order costs for cuisine types are positively skewed.
  - ◇ Exception: French- slight negative skew, & Thai- very slight negative skew.
- ⤴ Mediterranean cuisine has several outliers.
- ⤴ Korean & Vietnamese cuisine low order count inhibits interpretation.
- ⤴ Median cost around \$15 for most cuisine types.
  - ◇ Exception: Thai & French cuisines higher medians, around \$20.

## Cuisine type & prep time



- ⤴ High variability in median food preparation time for different cuisine types.
  - ◊ Low prep time cuisine types: Korean, Vietnamese, French, Southern.
    - Potential influence of low order count.
  - ◊ Excluding low points, median varies roughly around 27 minutes.
- ⤴ All cuisine types show a range of 22 to 33 minutes for the 50% of data points between the 25<sup>th</sup> and 75<sup>th</sup> quartile.
- ⤴ Most cuisine food preparation time total range is between 20 and 34 minutes.

## Day of week & delivery time



- ⤴ Median delivery time on the weekend is lower than the median delivery time on weekdays.
- ⤴ IQR for weekday delivery time is lower than the IQR of weekend delivery times.
  - ◇ Less variability in weekday delivery times.
  - ◇ More predictable delivery times on weekdays.
- ⤴ No outliers for delivery times.

## Highest revenue restaurants

Restaurant	Total revenue
Shake Shack	\$3,579.53
The Meatball Shop	\$2,145.21
Blue Ribbon Sushi	\$1,903.95
Blue Ribbon Fried Chicken	\$1,662.29
Parm	\$1,112.76
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

▲ Top 14 restaurants with highest revenue out of 178 restaurants.

▲ Shake Shack number 1:

◇ Revenue over \$3,500.

◇ +\$1,400 to next highest restaurant.

▲ Meatball shop number 2:

◇ +\$200 to next highest restaurant.

▲ All other restaurants:

◇ Less than \$200 different from next entry.

◇ Except Blue Ribbon Fried Chicken & Parm.

○ Blue Ribbon \$550 greater revenue.

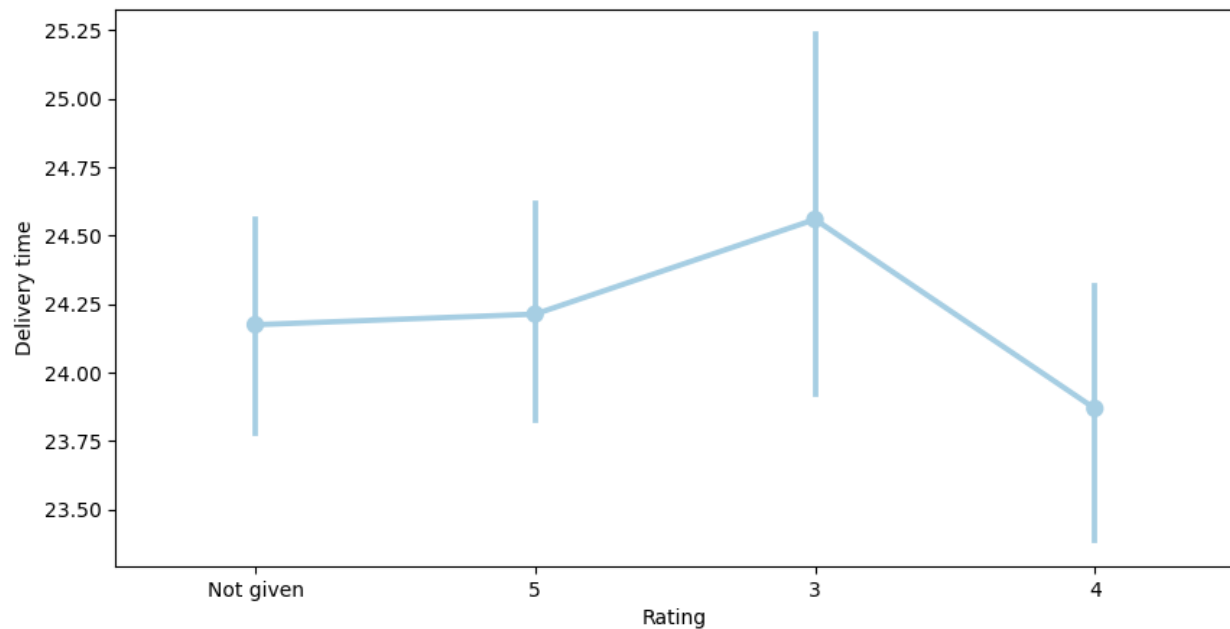
▲ Top 4 entries potential outliers.

▲ "Blue Ribbon Sushi" & "Blue Ribbon Sushi Bar & Grill" both listed.

◇ Worth checking if duplicate- not noted in dataset.

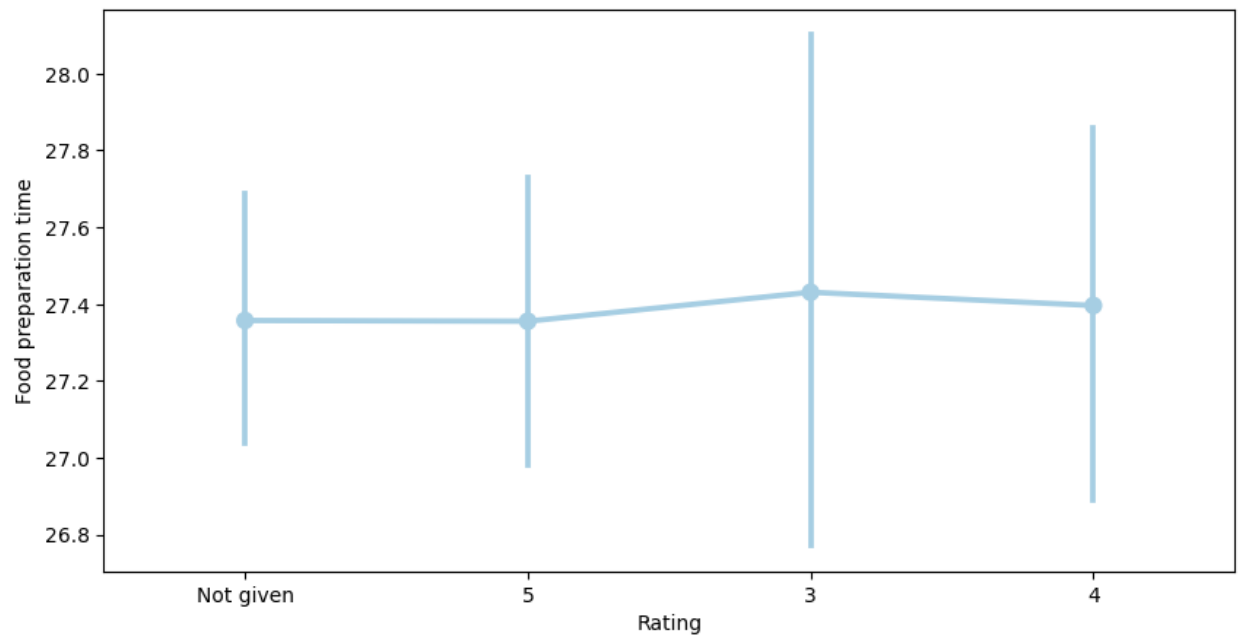


## Rating & delivery time



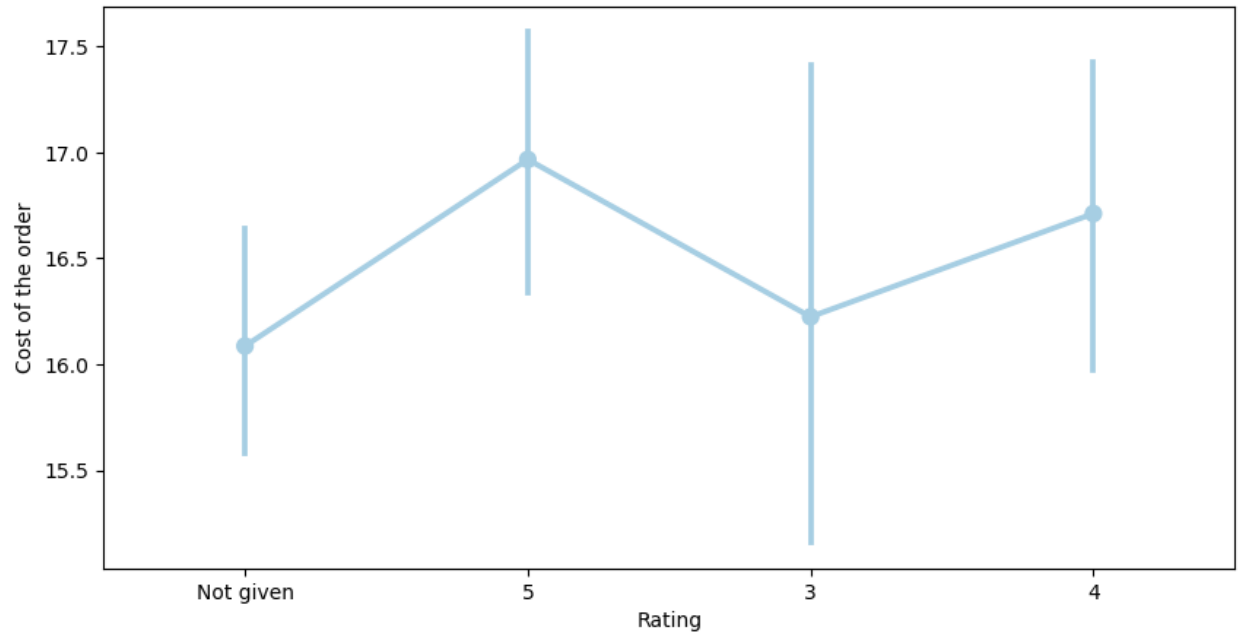
- ⤴ Orders with a rating of 3 are associated with the highest delivery times.
  - ◇ Roughly 24.5 minutes.
- ⤴ Orders with a rating of 4 are associated with the lowest delivery time.
  - ◇ Roughly 23.8 minutes.
- ⤴ Ratings of 5 and no rating are associated with delivery times that are neither the highest nor the lowest.
  - ◇ Roughly 24.2 minutes.

## Rating & food prep time



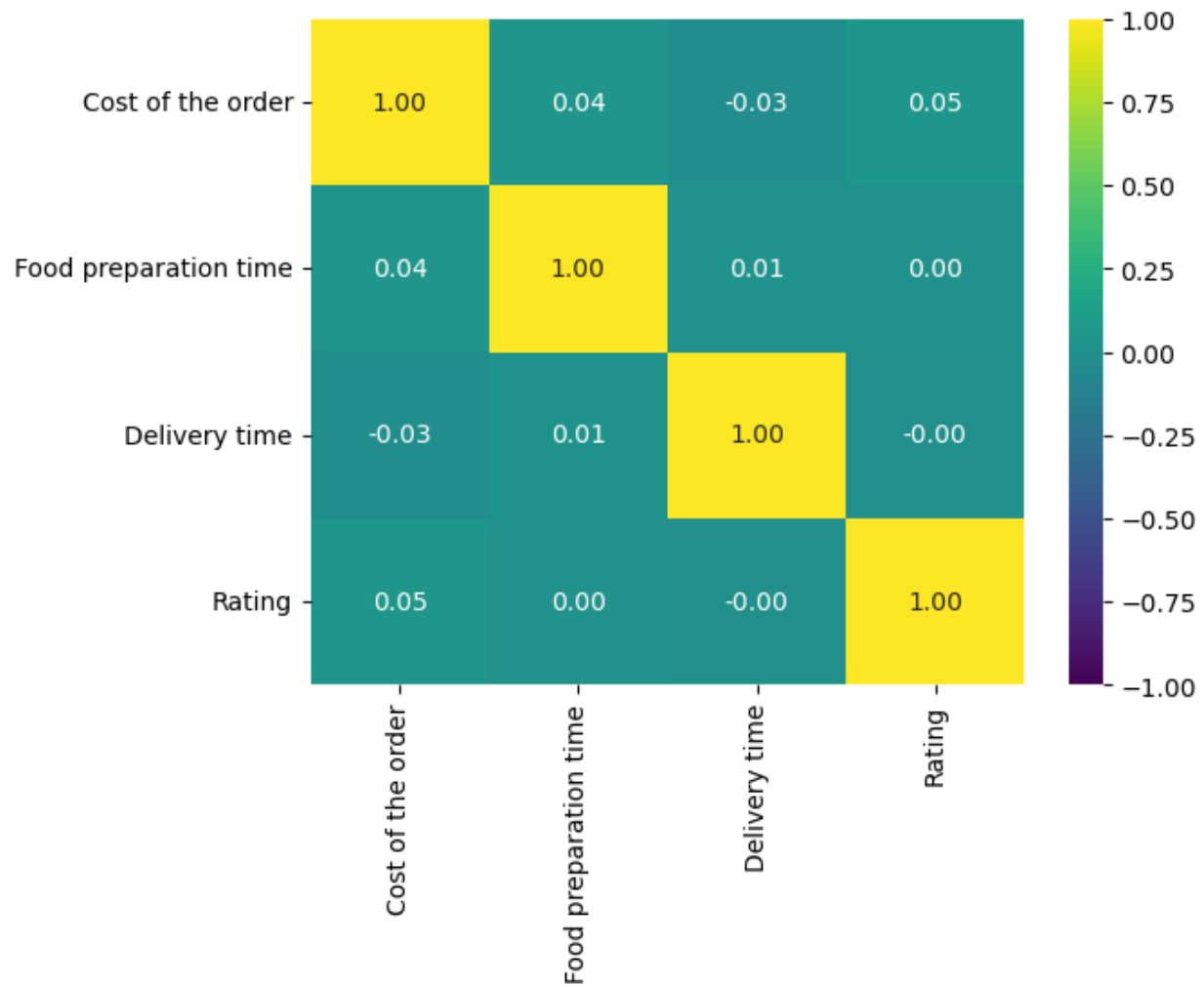
- ⤴ All ratings display roughly equivalent food preparation times.
  - ◇ Just under 27.4 minutes.

## Rating & order cost



- ▲ Order cost and rating appear to trend positively together.
  - ◇ Highest ratings (5) associated with highest order costs.
  - ◇ Next highest ratings (4) associated with next highest order costs.
  - ◇ Lowest ratings (3) associated with second lowest order costs.
  - ◇ No rating is associated with the lowest order cost.

## Variable correlations



- ⤴ Very little correlation between the variables of order cost, food prep time, delivery time, and rating.

## Promotional offer

- ⤴ Qualifying restaurants will have:
  - ◇ Over 50 ratings.
  - ◇ Average rating greater than 4.

### Restaurants with most ratings:

Top 5 restaurants	No. ratings
Shake Shack	133
The Meatball Shop	84
Blue Ribbon Sushi	73
Blue Ribbon Fried Chicken	64
RedFarm Broadway	41

### Qualifying restaurants:

Top 4 restaurants	Avg. rating
The Meatball Shop	4.511905
Blue Ribbon Fried Chicken	4.328125
Shake Shack	4.278195
Blue Ribbon Sushi	4.219178

- ⤴ 4 restaurants have over 50 ratings with an average rating higher than 4.0.
- ⤴ These restaurants qualify for the promotion.

## Net revenue

- ⤴ Revenue percentage by order cost:
  - ◇ 15% on orders over \$5.
  - ◇ 25% on orders over \$20.
- ⤴ Net revenue is roughly \$6,166.30.

## Delayed orders

- ⤴ Orders taking longer than 60 minutes:
  - ◇ Food preparation time and Delivery time combined.
- ⤴ Number of orders taking longer than 60 minutes to deliver: 200.
- ⤴ Percentage of orders taking longer than 60 minutes to deliver: 10.54%.

## Day of week delivery time

- ⤴ Orders on weekdays average around 28 minutes.
- ⤴ Orders on weekends average around 22 minutes.

## Conclusions and Recommendations

### Data quality

- Good size to dataset with robust variables.
- Nearly 40% of orders not rated.
  - ◇ Potential to increase for better understanding of customer experience.
  - ◇ Unrated orders associated with low-cost orders.
    - Explore why customers feel better incentivized to rate when order costs are higher.

## Univariate analysis

- American cuisine strongly positioned as most popular cuisine type.
  - ◇ Overall & on weekends at peak order time.
  - ◇ Can focus approaches with incentives or revenue on this cuisine type to target a large proportion of orders.
- Order costs tend to be less than the median cost, with less than 30% of the orders costing more than \$20.
  - ◇ Target order cost increase for increased revenue.
    - Information needed on drivers behind customer order cost.
- Orders more frequent on weekends in comparison to weekdays.
  - ◇ Determine strategies to increase weekday orders.
  - ◇ Target weekend order revenue for increase on a high proportion of orders.
- Food preparation times are generally consistent, but food delivery times tend to be greater than the median.
  - ◇ The total length of customer wait time is more greatly affected by delivery time.



→ Shake Shack top performing restaurant by number of orders.

- ◇ High degree of difference between Shake Shack and next-best performing restaurant.
- ◇ More information on restaurant choice drivers can lead to better insight into relative frequency of orders between restaurants.
  - Influence of marketing strategies by restaurants.
  - Total delivery time & restaurant choice.
  - Other factors.
- ◇ Can use insight on order frequency to target increased quantity of orders overall, or to allocate more sales to higher order cost restaurants.

→ 20% discount voucher to top 3 customers by number of orders may increase customer retention.

- ◇ Represents a very small proportion of the number of orders.
  - Little revenue loss.
  - Effect should be evaluated for overall customer satisfaction.

## Multivariate analysis

→ The type of cuisine of an order does not have a strong effect on the minimum order cost, but it does have an effect on the distribution of order cost.

- ◇ While the order cost for most cuisine types tends to fall below the median, French & Vietnamese cuisines tend to incur order costs above the median.

- Target these cuisine types for better understanding of relationship between menu items, customer order drivers, and total order cost.

→ Weekend delivery times are shorter but more variable than weekday times.

- ◇ If more consistent delivery times are needed, focus on weekend.

- ◇ If shorter delivery times are needed, focus on weekday.

→ Shake shack top performing restaurant by revenue.

- ◇ High number of orders is likely the biggest contributor.

- ◇ \$1,400 increase in revenue over next highest indicates Shake Shack performance is a good target for future analysis and insight.

→ Rating values appear to vary the most strongly with order cost and delivery time, but correlation values were low in magnitude for both variables.

- ◇ Other variables on customer satisfaction may be a better indicator of customer satisfaction.

- ◇ Recommend analyzing order cost and delivery time with another customer satisfaction metric or improving rating frequency for more accurate data.

→ Promotional incentive will accelerate sales on an already well-performing sector of the restaurants.

- ◇ Italian, American, & Japanese restaurants represented in promotional offer.

- 3 cuisine types representing the top 3 sold by number of orders.