FoodHub Business Analysis

Project Week 5
AIML - UT Austin

Prepared by Greg Wenzel

June 21, 2023

Contents / Agenda

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

Executive Summary

Analysis of the food delivery data from FoodHub provides valuable insights to enhance the customer experience and improve the business. We have addressed key questions and identified areas of opportunity for FoodHub.
Here are the key findings:

- Demand:

Data was analyzed to understand the demand for different restaurants and cuisines.
We identified the top revenue-generating restaurants which can be a focus for marketing & partnerships.
We determined the most popular cuisine, which can guide the company in expanding & targeting specific customer segments.

- Customer Satisfaction:

Customer ratings play a crucial role in determining customer satisfaction.
We examined the average ratings given by customers and identified the restaurants with the highest & lowest ratings.
Improving the ratings of low-rated restaurants can help enhance customer satisfaction and loyalty.

- Operational Efficiency:

We assessed the time taken for food preparation and delivery.
By analyzing food preparation & delivery time, we can identify areas for improvement & optimize the delivery process.
Identifying restaurants with longer preparation times can help FoodHub work closely with them to reduce the waiting time for customers.

- Weekday vs. Weekend:

We analyzed the differences in customer behavior and demand between weekdays and weekends.
Understanding the variations in order volume, revenue, and customer preferences on different days of the week can assist in strategic decision-making and resource allocation.

Executive Summary

Recommendations: We recommend the following actions to enhance the business:

- Strengthen Partnerships:

Strengthen partnerships with top revenue-generating restaurants to improve their visibility on the app and attract more customers.

Explore collaboration opportunities with popular restaurants to expand the cuisine offerings and cater to diverse customer preferences.

- Improve Customer Satisfaction:

Work closely with low-rated restaurants to address customer concerns and improve their ratings.

Implement a feedback loop to gather customer feedback and proactively address any issues raised.

- Optimize Operations:

Collaborate with restaurants experiencing longer food preparation times to streamline their processes and reduce waiting time.

Enhance coordination with delivery personnel to improve delivery time and ensure timely and efficient service.

- Targeted Marketing:

Leverage the insights gained from the weekday vs. weekend analysis to tailor marketing strategies for specific days.

Promote offers and discounts during weekdays to boost demand during slower periods.

By implementing these recommendations, FoodHub can enhance the customer experience, improve operational efficiency, and drive business growth. Continuous monitoring and analysis of customer data will provide further insights for ongoing optimization.

Data and marketing teams can use these findings to make informed decisions and drive impactful actions that benefit both the customers and the business.

Question 17 is Referenced Here!

Business Problem Overview and Solution Approach

Problem Statement:

- The food aggregator company, FoodHub, aims to analyze the data of different food orders made by registered customers in their online portal. The objective is to gain insights into the demand for different restaurants and cuisines to enhance the customer experience and improve the overall business.

Key Questions:

- What are the top revenue-generating restaurants?
- Which cuisines are most popular among customers?
- How can customer satisfaction be improved based on ratings?
- Are there opportunities to optimize food preparation and delivery times?
- What are the variations in customer behavior and demand between weekdays and weekends?

Data Description: The dataset includes various details related to food orders, such as order ID, customer ID, restaurant name, cuisine type, cost of the order, day of the week, customer rating, food preparation time, and delivery time.

Objective: Perform data analysis on the provided data to address the key questions and provide insights and recommendations to FoodHub. By understanding the demand patterns, customer preferences, and operational efficiencies, FoodHub can make informed decisions to enhance their customer experience, strengthen partnerships with top restaurants, optimize operations, and strategically target marketing efforts.

The analysis will contribute to improving the overall business performance of FoodHub and ensuring they remain a preferred choice for customers seeking online food delivery services in a competitive market.

Business Problem Overview and Solution Approach

Solution Approach / Methodology:

- To address the problem statement and answer the key questions, the following solution approach and methodology can be applied:

Data Understanding:

- Start by gaining a clear understanding of the provided dataset, including its structure, variables, and data dictionary.
- Perform exploratory data analysis (EDA) to identify any data quality issues, missing values, or outliers that may affect the analysis.

Data Preprocessing and Cleaning:

- Clean the data by handling missing values, outliers, and any inconsistencies in the data.
- Convert data types if necessary and ensure data is in the appropriate format for analysis.

Descriptive Analysis:

- Conduct descriptive analysis to explore the dataset and understand the distribution of variables.
- Calculate basic statistical measures such as mean, median, mode, and standard deviation to gain insights into the data.

Demand Analysis:

- Analyze the demand for different restaurants by examining the number of orders and revenue generated.
- Identify the top revenue-generating restaurants and popular cuisine types among customers.

Customer Satisfaction Analysis:

- Analyze customer ratings to understand satisfaction levels.
- Identify restaurants with the highest and lowest ratings to address any issues and improve customer satisfaction.

Business Problem Overview and Solution Approach

Operational Efficiency Analysis:

- Analyze food preparation time and delivery time to identify areas for improvement and operational optimization.
- Identify restaurants with longer preparation times and work closely with them to reduce waiting time for customers.

Weekday vs. Weekend Analysis:

- Analyze variations in customer behavior, order volume, revenue, and customer preferences between weekdays and weekends.
- Identify trends and patterns to make informed decisions regarding resource allocation and marketing strategies.

- Visualization and Reporting:
- Utilize appropriate visualizations such as bar plots, line plots, box plots, heatmaps, etc., to present key findings.
- Prepare a comprehensive report summarizing, analysis, insights, & recommendations for the data scientist & marketing team.

- Recommendations:
- Provide actionable recommendations based on the analysis to enhance customer experience, improve customer satisfaction, optimize operations, and drive business growth.
- Focus on strengthening partnerships with top revenue-generating restaurants, expanding cuisine offerings, addressing low-rated restaurants' concerns, and targeted marketing strategies.

- Continuous Monitoring and Iterative Analysis:
- Implement regular monitoring of key metrics and repeat the analysis periodically to track progress and identify new opportunities for improvement.

By following this solution approach, the data and marketing teams can gain valuable insights, make data-driven decisions, and take appropriate actions to enhance the customer experience and drive the success of FoodHub.

Data Overview

Data Description :
The dataset includes various details related to food orders:

- **order ID,**
- **customer ID,**
- **restaurant name**
- **cuisine type**
- **cost of the order**
- **day of the week**
- **customer rating**
- **food preparation time**
- **and delivery time.**

```
RangeIndex: 1898 entries, 0 to 1897
Data columns (total 9 columns):
 #   Column                 Non-Null Count  Dtype
---  ------                 --------------  -----
 0   order_id               1898 non-null   int64
 1   customer_id            1898 non-null   int64
 2   restaurant_name        1898 non-null   object
 3   cuisine_type           1898 non-null   object
 4   cost_of_the_order      1898 non-null   float64
 5   day_of_the_week        1898 non-null   object
 6   rating                 1898 non-null   object
 7   food_preparation_time  1898 non-null   int64
 8   delivery_time          1898 non-null   int64
```

Data Overview

Data related questions:

1. How many rows and columns in Data : 1889 Rows, 9 Columns
2. What are the datatypes of the columns: Integers, Objects and Floating Integers
3. Are there missing values in the data : No
4. What is the minimum, average & maximum time for consumer to receive the delivery:

- Average : 51.53 minutes
- Maximum : 68 Minutes
- Minimum: 35 minutes

5. How many orders are not rated : 736

```
There are no missing values in the data:
order_id                0
customer_id             0
restaurant_name         0
cuisine_type            0
cost_of_the_order       0
day_of_the_week         0
rating                  0
food_preparation_time   0
delivery_time           0
dtype: int64
```

```
Statistical Summary of Numerical Data
            order_id    customer_id  cost_of_the_order  food_preparation_time
count   1.898000e+03    1898.000000        1898.000000            1898.000000
mean    1.477496e+06  171168.478398          16.498851              27.371970
std     5.480497e+02  113698.139743           7.483812               4.632481
min     1.476547e+06    1311.000000           4.470000              20.000000
25%     1.477021e+06   77787.750000          12.080000              23.000000
50%     1.477496e+06  128600.000000          14.140000              27.000000
75%     1.477970e+06  270525.000000          22.297500              31.000000
max     1.478444e+06  405334.000000          35.410000              35.000000

       delivery_time
count    1898.000000
mean       24.161749
std         4.972637
min        15.000000
25%        20.000000
50%        25.000000
75%        28.000000
max        33.000000
```

Univariate Analysis

Question 6: Explore all the variables and provide observations on their distributions.

Basic observations of single variants:

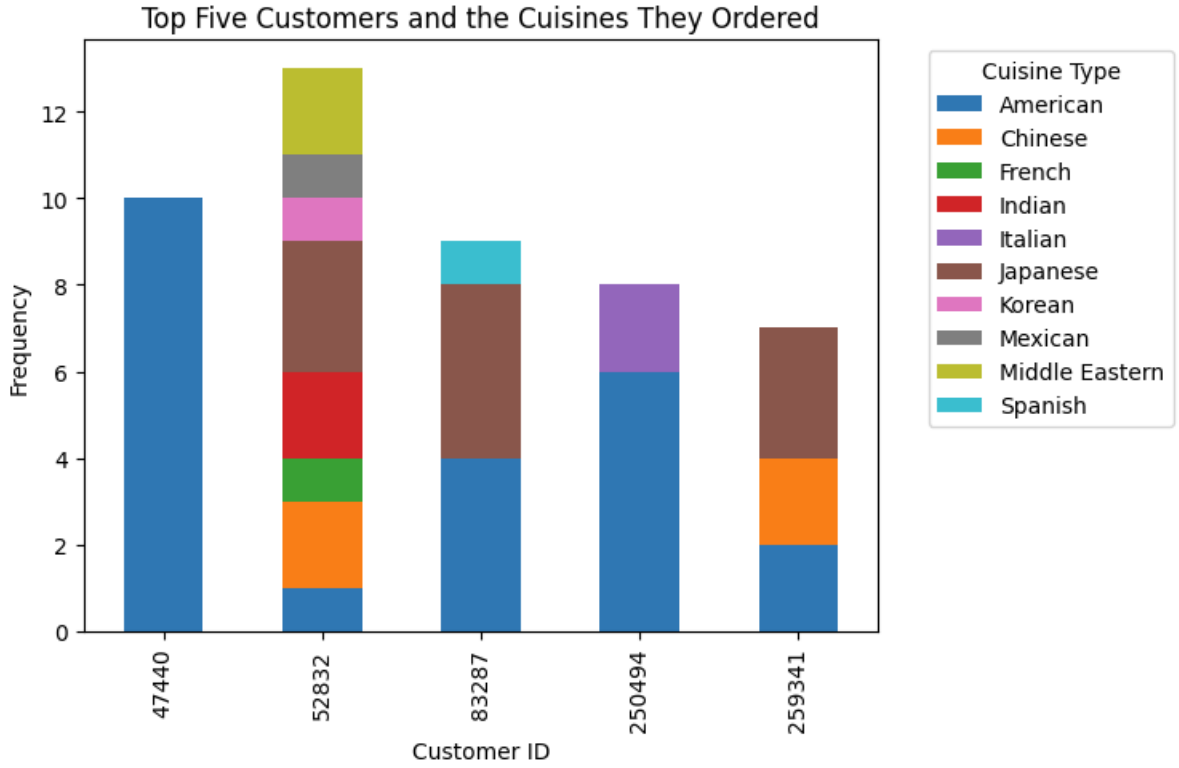
- 1200 unique customers
- 178 Restaurants
- 14 different cuisines
- American cuisine = most ordered at 31%
- Vietnamese = least ordered at less than ½ of a percent
- Average cost per order = $16.50
- 71% of the orders delivered on the weekend
- 29% of the orders delivered on the weekdays
- French food = most expensive averaging over $19.00
- Vietnamese = least expensive averaging just under $13.00
- American cuisine averaged $16.32 compared to the group average of $16.50
- 29.24% of orders cost above $20 numbering 555
- Average delivery time for all orders was 24.16 minutes (restaurant pickup until delivered)
- Average delivery time from order placement until delivery = 51.53 minutes

Variables:
Order ID
Customer ID
Restaurants
Cuisine Types
Cost of Order
Weekend / Weekday
Rating
Preparation Time
Delivery Time

Univariate Analysis Continued

Repeat Customers = 58.17%, Highest Frequency = 13, Total Repeat = 698

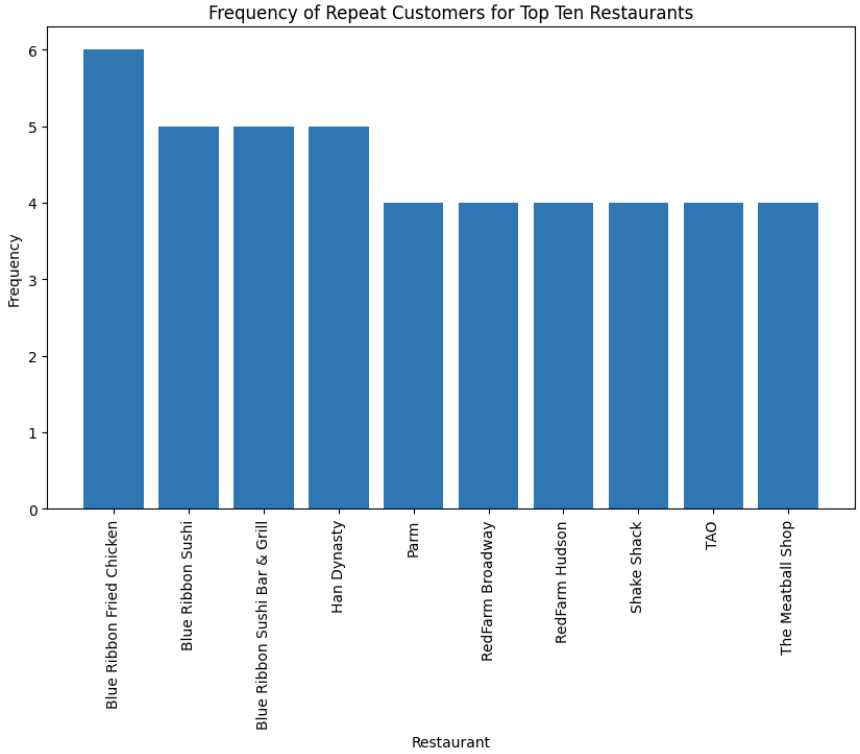
Question 6: ***Additional Observations***



Univariate Analysis Continued

Blue Ribbon Fried Chicken has the highest number of repeat customers!

Question 6: ***Additional Observations***

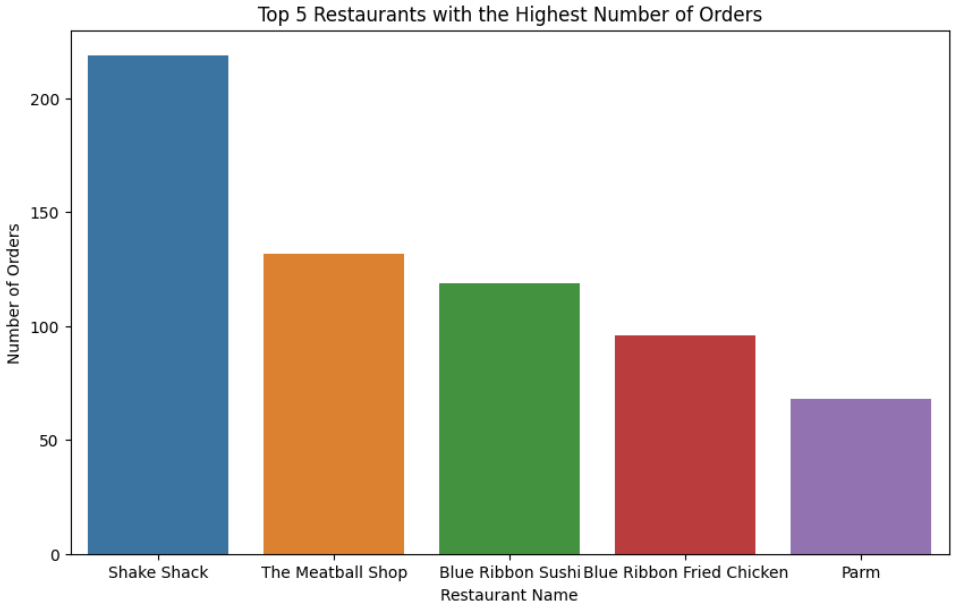


Univariate Analysis Continued

Question 7: *Which are the top 5 restaurants in terms of the number of orders received?*

Top 5 restaurants with the highest order count:

1. Shake Shack : 219
2. The Meatball Shop : 132
3. Blue Ribbon Sushi : 119
4. Blue Ribbon Fried Chicken : 96
5. Parm : 68



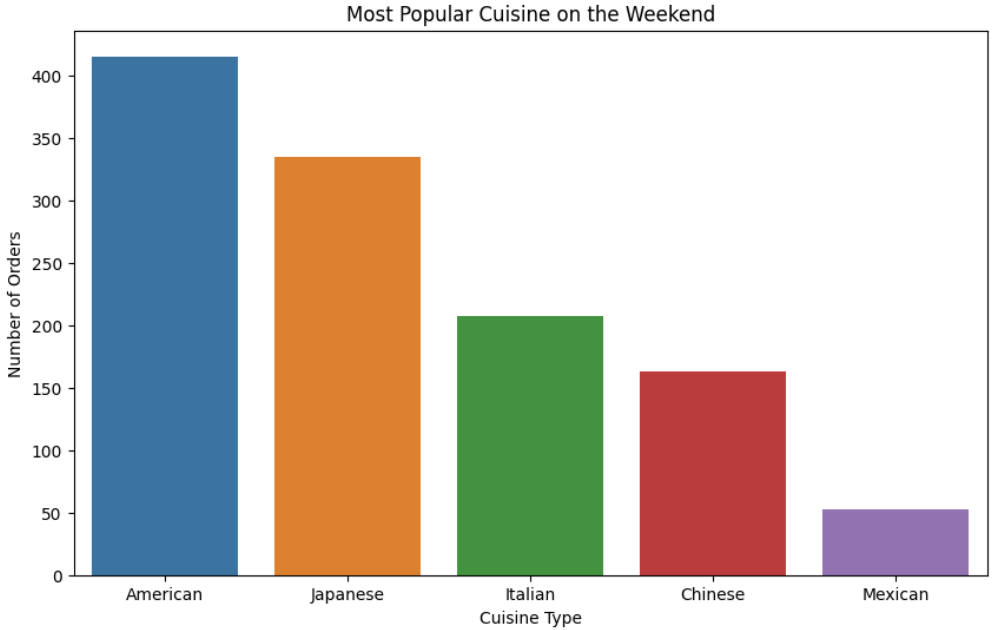
Univariate Analysis Continued

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

Most popular cuisine on the weekend is American

Note: American cuisine is the most popular cuisine regardless of day

1. American 415
2. Japanese 335
3. Italian 207
4. Chinese 63
5. Mexican 53



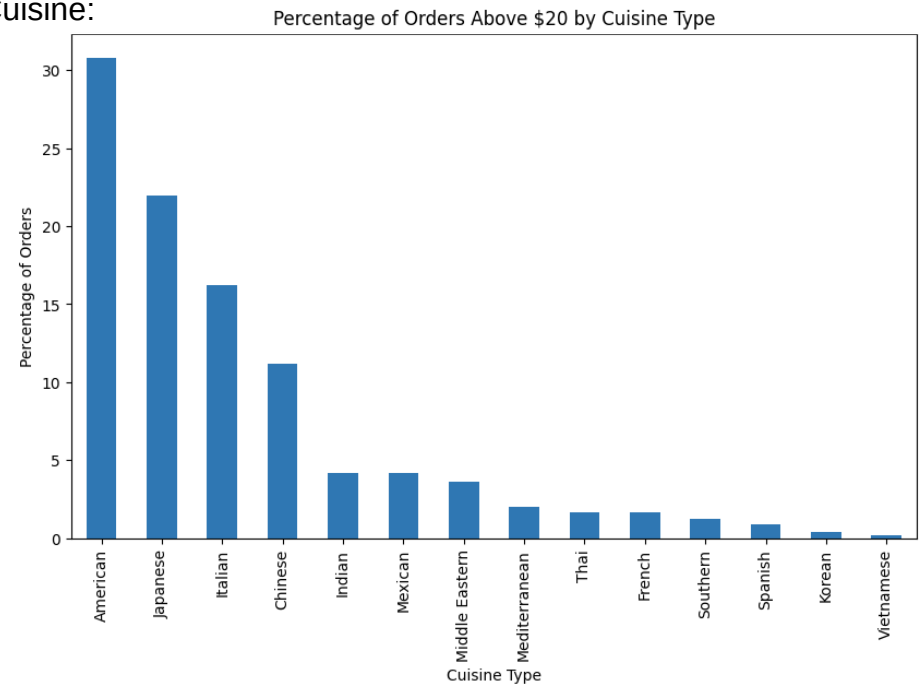
Univariate Analysis Continued

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

Percentage of orders above 20 dollars: 29.24 % = Total of 555 Orders

Percentages of orders above $20 by Cuisine:

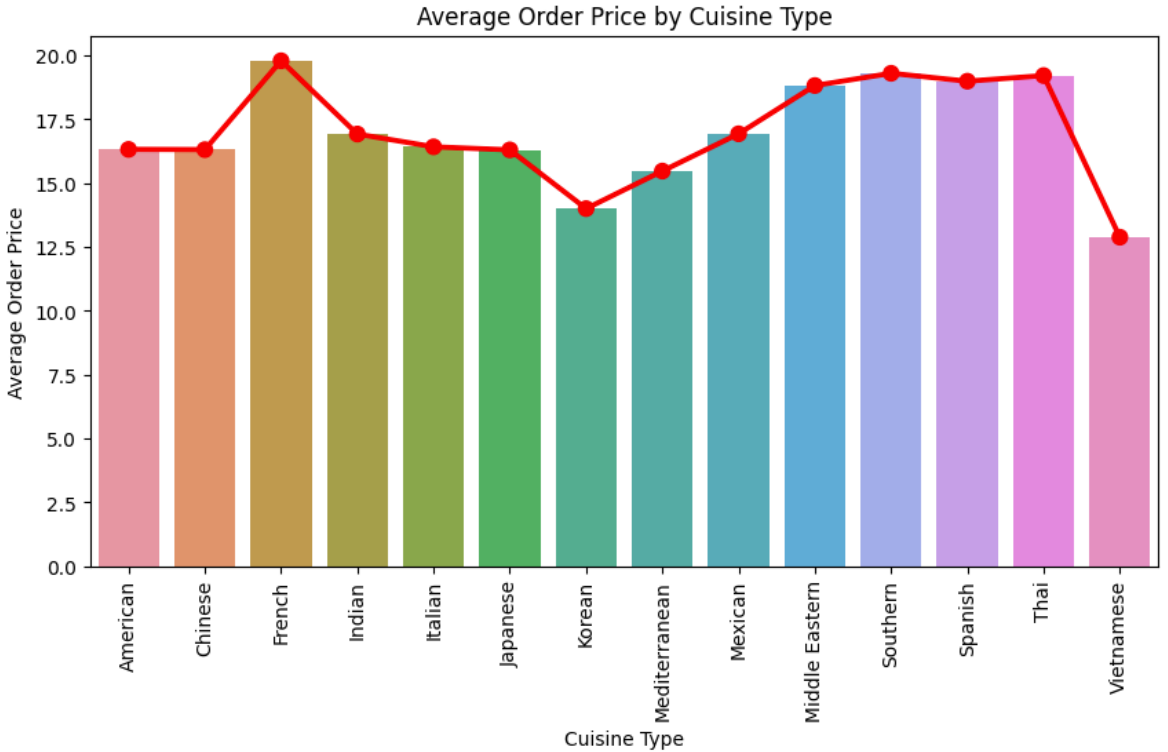
1. American 30.81
2. Japanese 21.98
3. Italian 16.22
4. Chinese 11.17
5. Indian 4.14
6. Mexican 4.14
7. Middle Eastern 3.60
8. Mediterranean 1.98
9. Thai 1.62
10. French 1.62
11. Southern 1.26
12. Spanish 0.90
13. Korean 0.36
14. Vietnamese 0.18



Univariate Analysis Continued

Question 9: Additional Observations

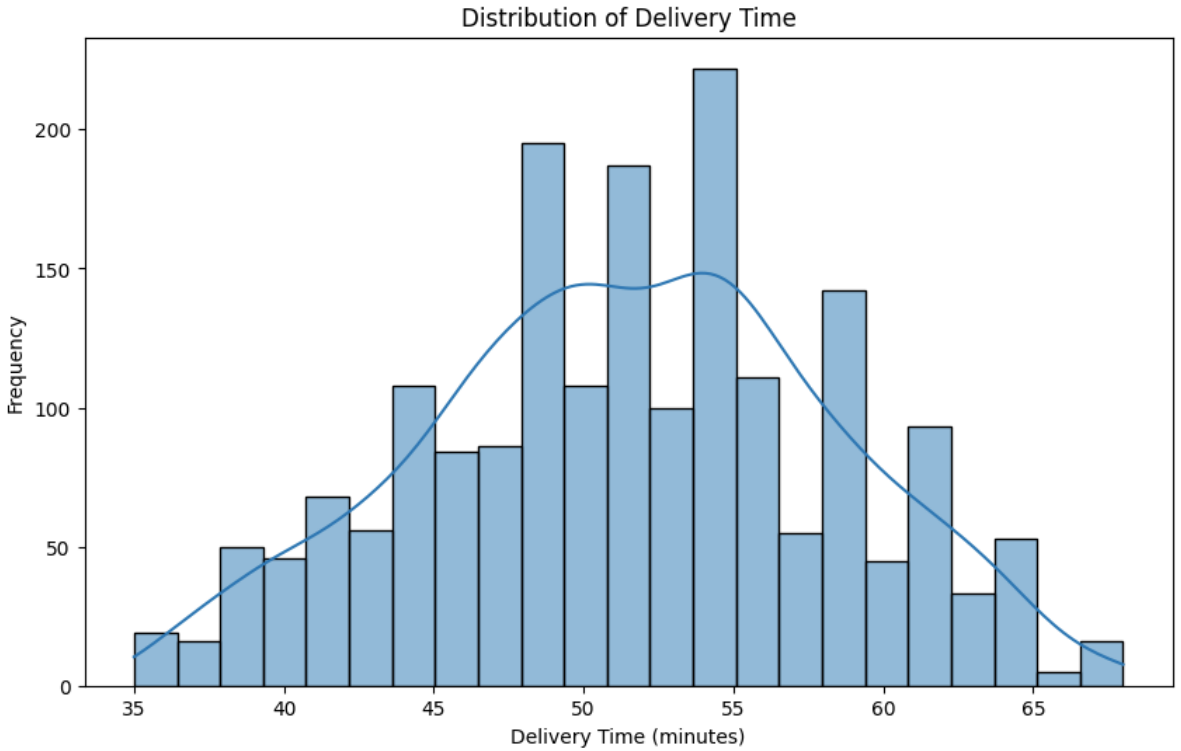
```
Mean Order Price per Cuisine Type:
cuisine_type
American          16.32
Chinese           16.31
French            19.79
Indian            16.92
Italian           16.42
Japanese          16.30
Korean            14.00
Mediterranean     15.47
Mexican           16.93
Middle Eastern    18.82
Southern          19.30
Spanish           18.99
Thai              19.21
Vietnamese        12.88
```



Univariate Analysis Continued

Question 10: What is the mean order delivery time? **51.53** Minutes *(ordered until delivered)*

```
The mean delivery time for this dataset is 51.53 minutes
```

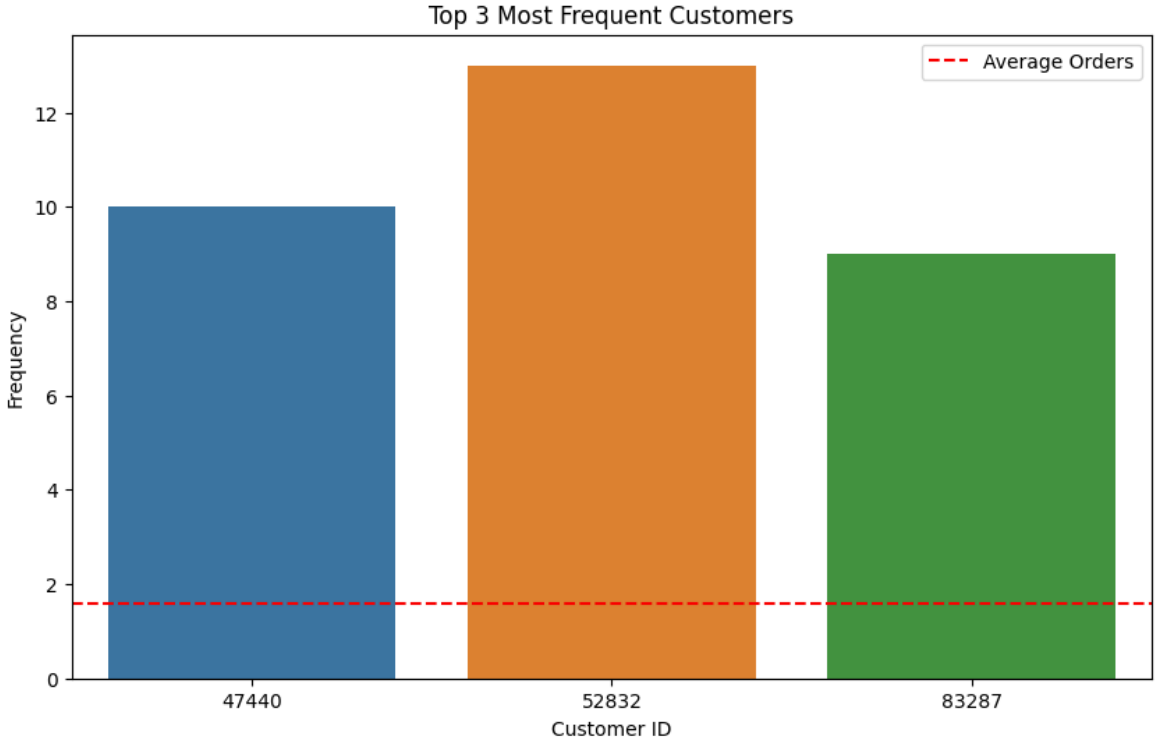


Univariate Analysis Continued

Question 11: The company has decided to give 20% discount vouchers to the top 3 most frequent customers. Find the IDs of these customers and the number of orders they placed.

Average number of orders for all customers is 1.57

Customer 52832 : Orders = 13
Customer 47440 : Orders = 10
Customer 83287 : Orders = 9



Multivariate Analysis

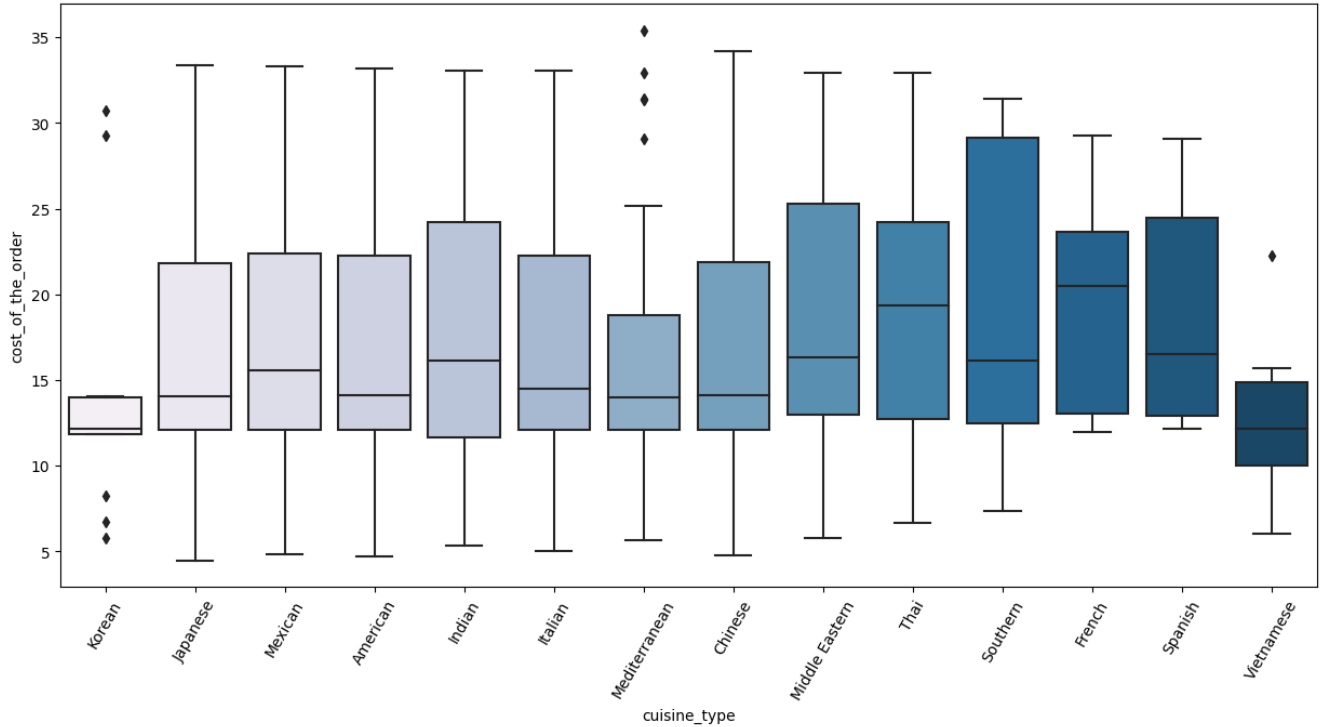
- In this section of the analysis we will explore multiple variables simultaneously to gain insights into how different variables interact with each other and make a collection of contributions to the patterns analyzed in the data.
- We will:

1. Identify relationships within the data
2. Gain understanding of patterns and how they relate to other areas of the data
3. Implement code libraries and techniques to visualize multidimensional data
4. Predict outcomes analyzing multiple variables & their influence on each other & the data

Multivariate Analysis

Question 12: Perform a multivariate analysis to explore relationships between the important variables in the dataset.

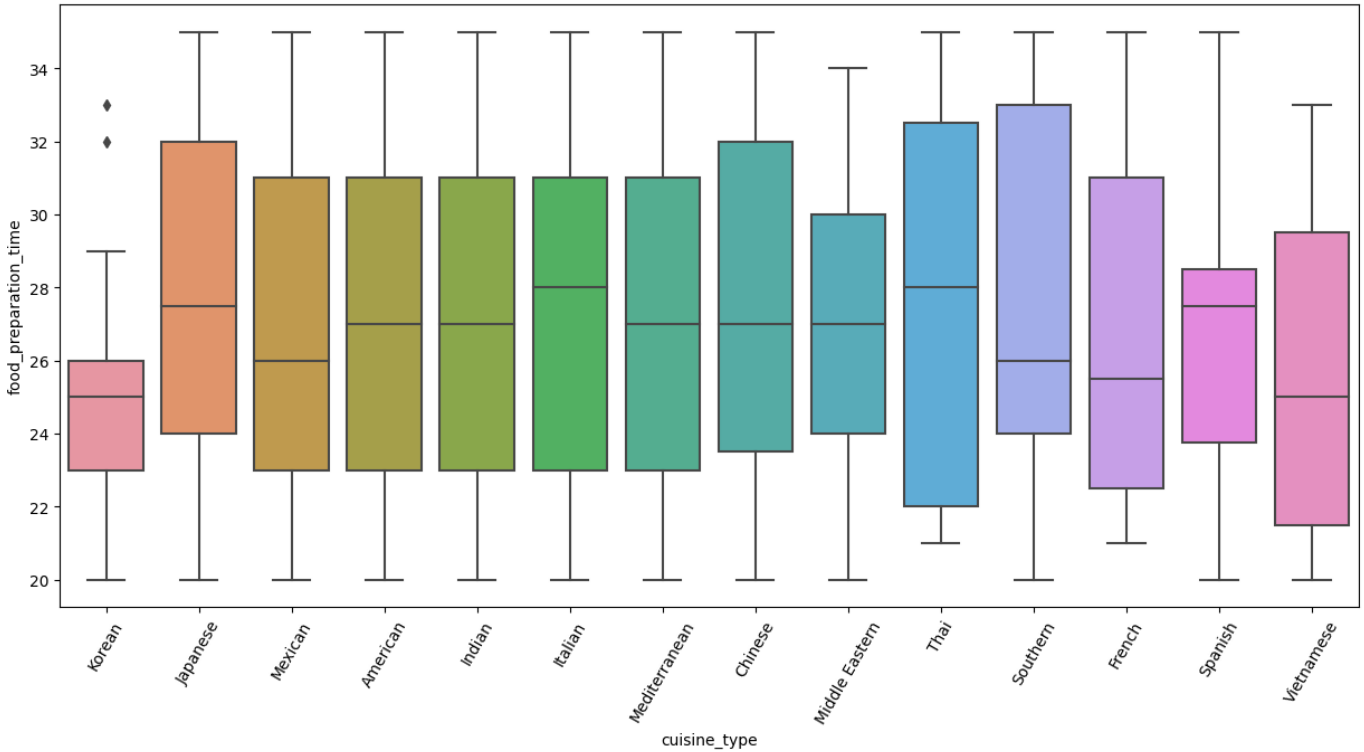
- ***Cuisine vs Cost of the order***



Multivariate Analysis Continued

Question 12:

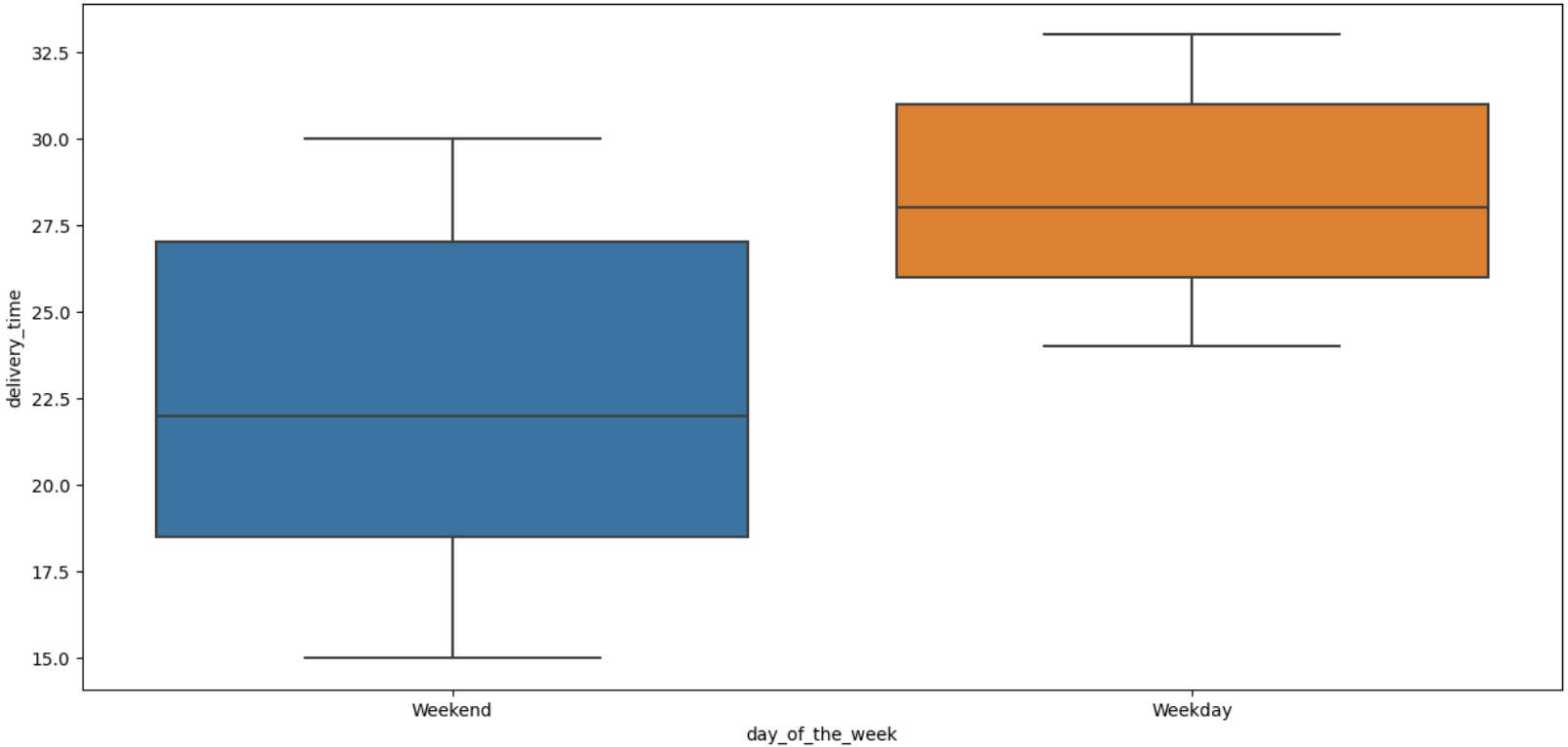
- **Cuisine vs Food Preparation Time**



Multivariate Analysis Continued

Question 12:

- **Day of the Week vs Delivery time**

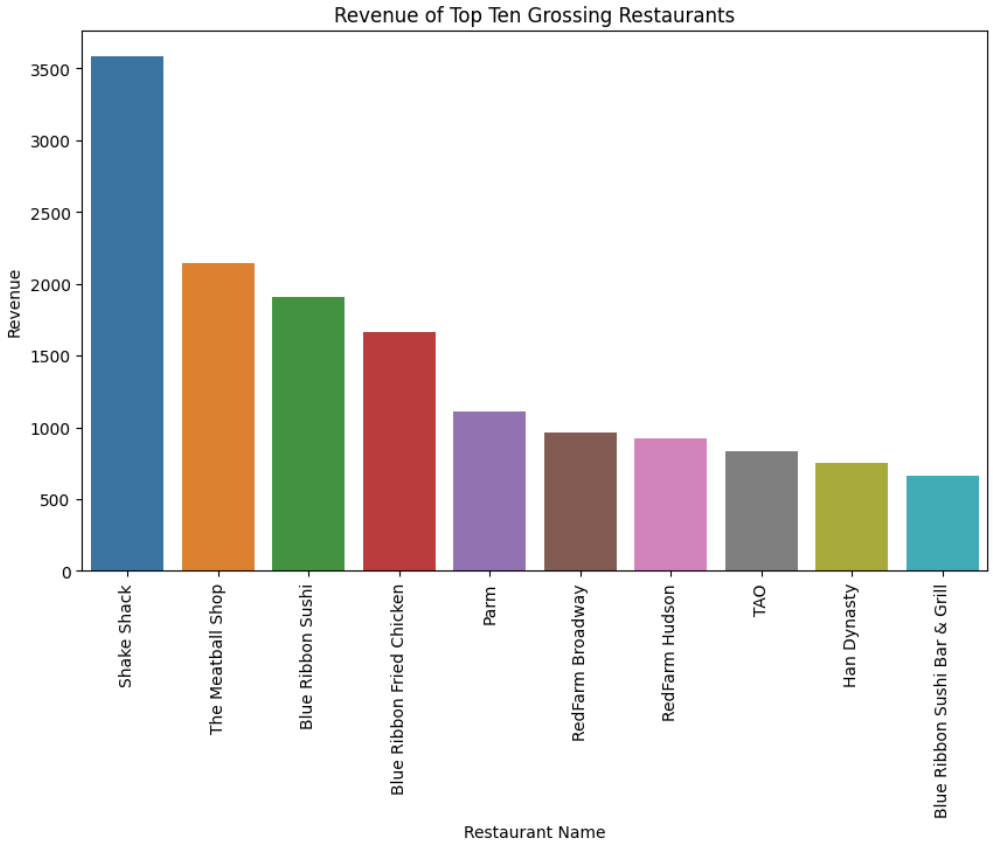


Multivariate Analysis Continued

Question 12: Observations regarding revenue of top ten grossing restaurants

The top ten grossing restaurants are

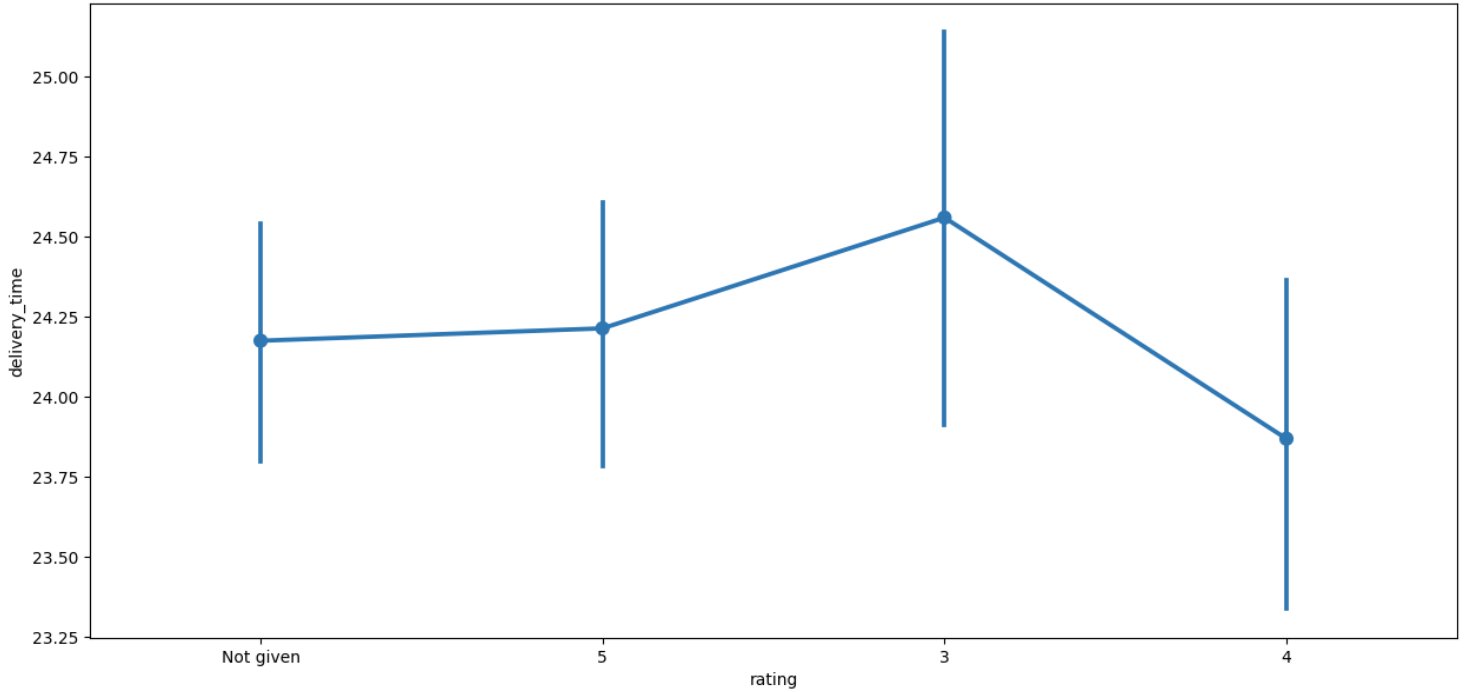
restaurant_name	
Shake Shack	3579.53
The Meatball Shop	2145.21
Blue Ribbon Sushi	1903.95
Blue Ribbon Fried Chicken	1662.29
Parm	1112.76
RedFarm Broadway	965.13
RedFarm Hudson	921.21
TAO	834.50
Han Dynasty	755.29
Blue Ribbon Sushi Bar & Grill	666.62



Multivariate Analysis Continued

Question 12: This graph represents the time from pickup at restaurant to delivery

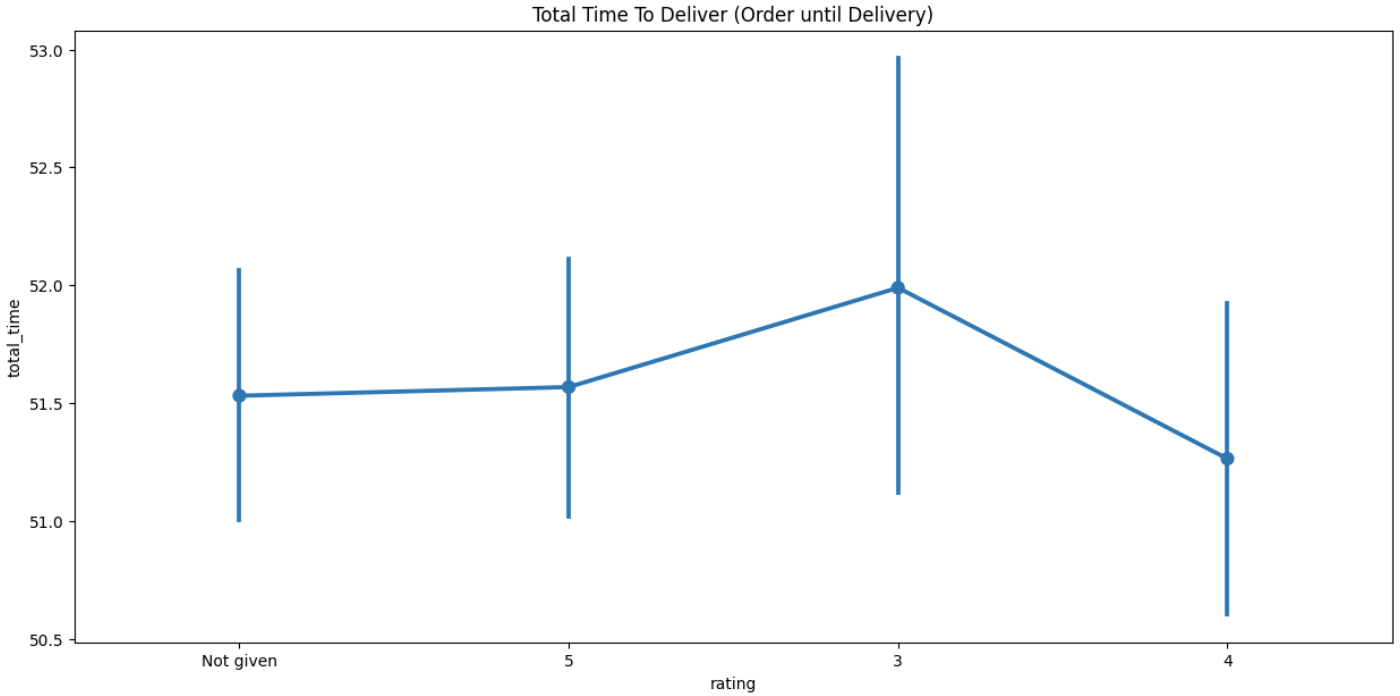
- Rating vs Delivery Time



Multivariate Analysis Continued

Question 12: This graph represents the time the order was placed until delivered

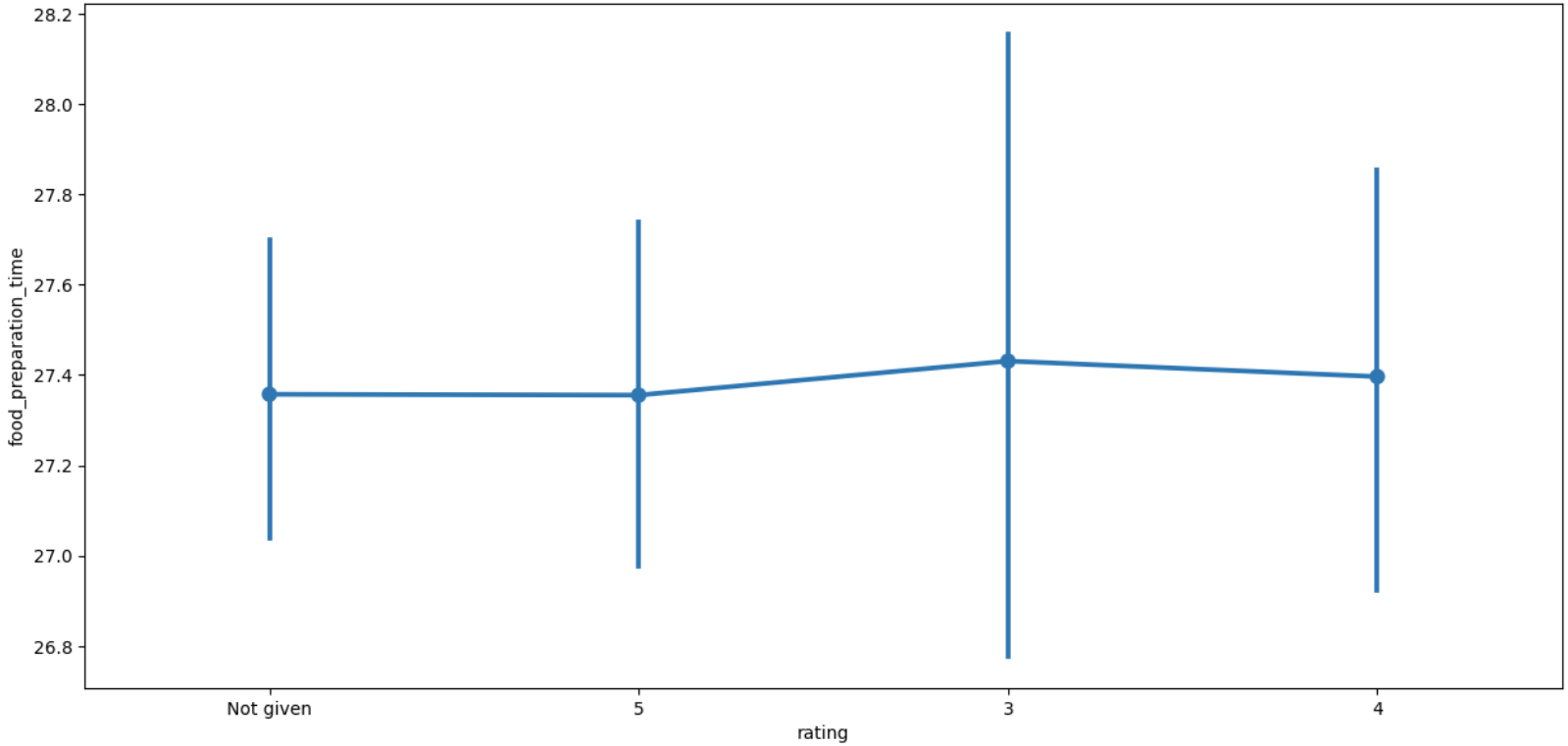
- Rating vs Delivery Time #2



Multivariate Analysis Continued

Question 12:

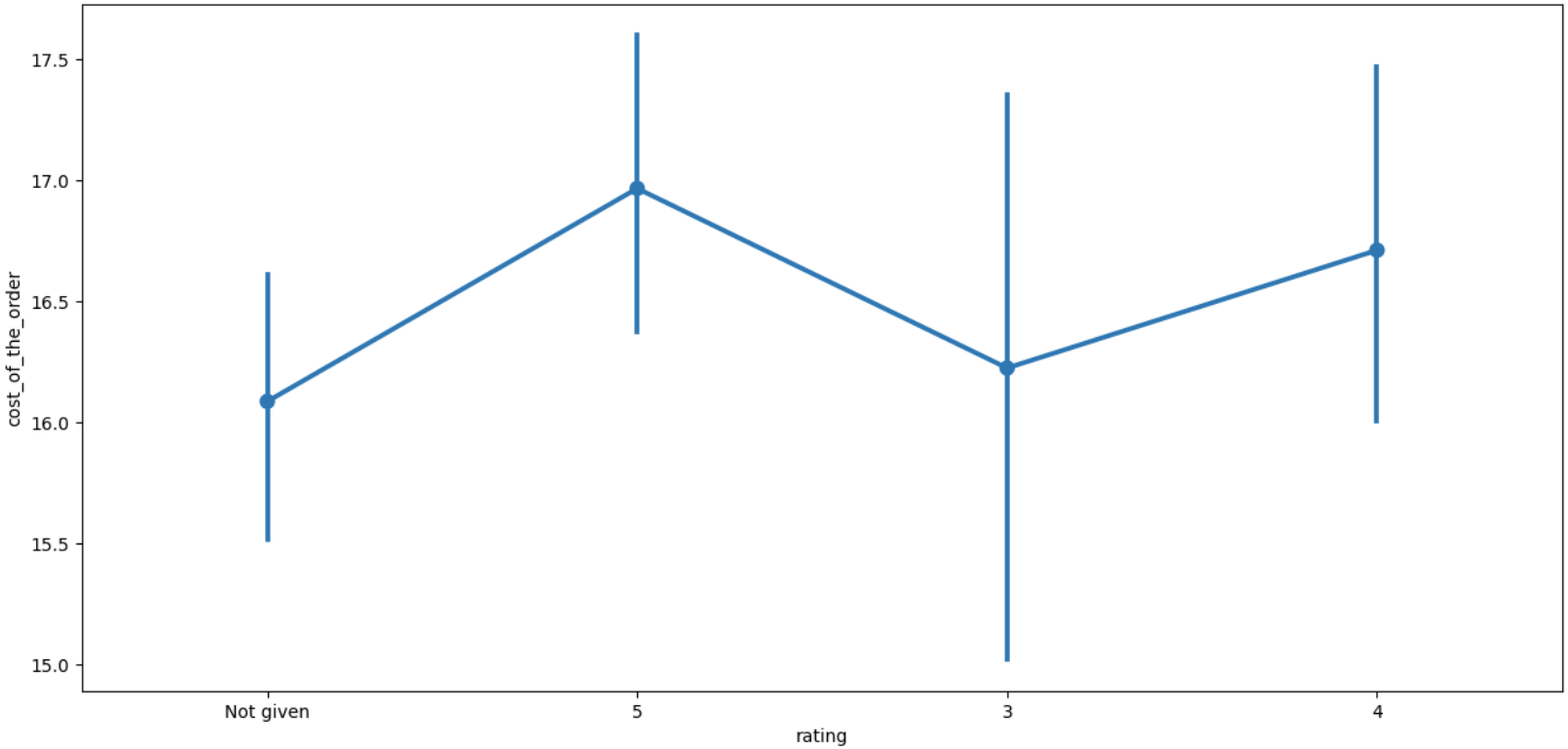
- **Rating vs Food Preparation Time**



Multivariate Analysis Continued

Question 12:

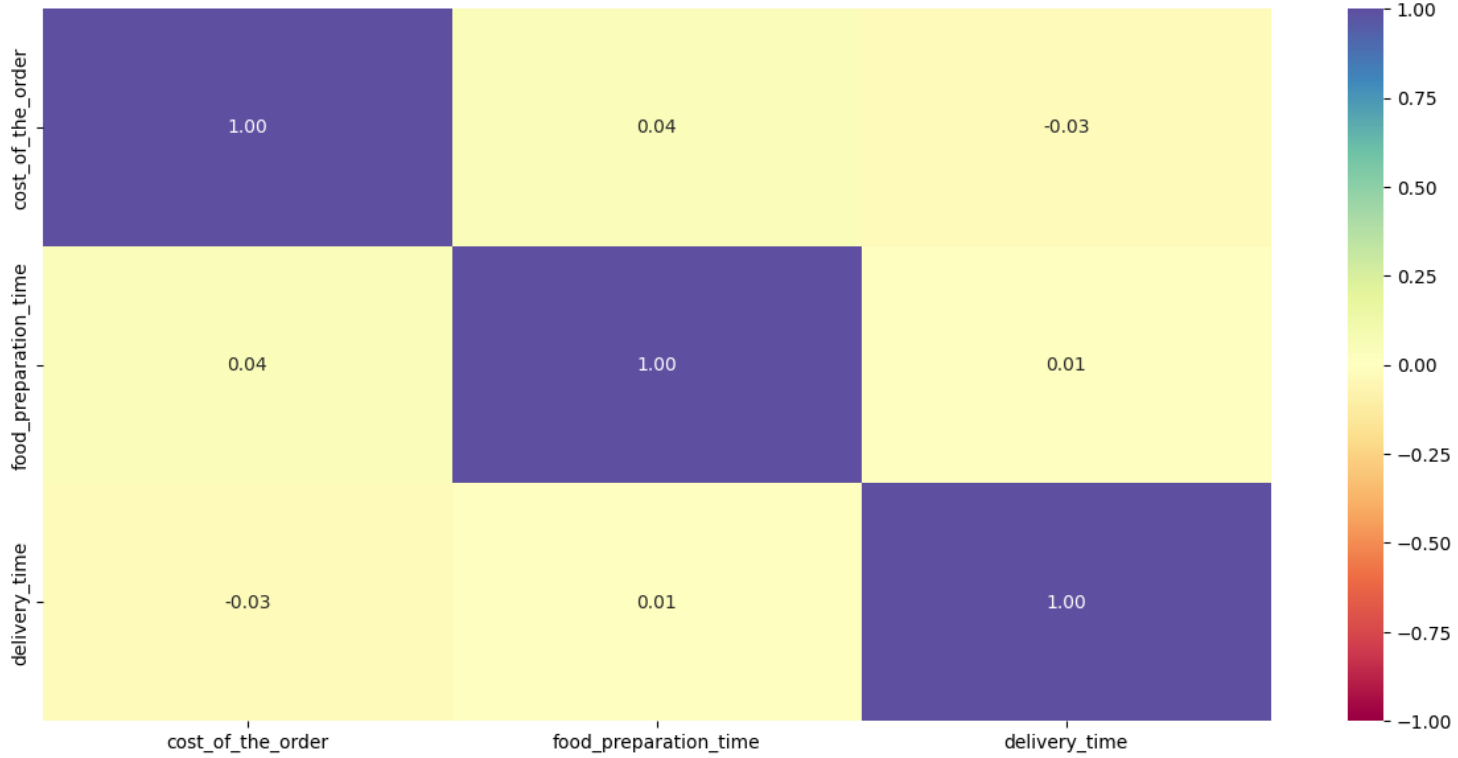
- Rating vs Cost of the Order



Multivariate Analysis Continued

Question 12:

- **Correlation Among Variables**



Multivariate Analysis Continued

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.

Restaurant Rankings Count

	restaurant_name	rating
0	Shake Shack	133
1	The Meatball Shop	84
2	Blue Ribbon Sushi	73
3	Blue Ribbon Fried Chicken	64
4	RedFarm Broadway	41

Restaurant Rating

This group meets the criteria for promotion

	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 Continued

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. The net revenue is around $6166.30

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 percentage of orders over 60 minutes is 10.54
- The total number of orders over 60 minutes is 200
- The average time for tickets over 60 minutes is 51.53
- The total number of deliveries under 60 minutes is 1653
- The total number of orders is 1898

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 56 minutes
- The mean delivery time on weekends is around 50 minutes

Multivariate Analysis Continued

Question 17: ***Conclusions and Recommendations***

Conclusion:

- American Cuisine accounts for 31% of food orders & Vietnamese less than 1%
- Shake Shack has the highest gross and number of orders in the data = $3579.53 | 219
- French cuisine has the highest average cost while Vietnamese has the lowest.
- Southern Cuisine has the greatest distribution of cost while Korean has the lowest.
- The top ten grossing restaurants have between 6 and 4 repeat customers
- The largest grouping for cost of orders is around $12 at around 350 orders
- Weekend orders more than double the number of weekday orders
- Weekday orders average around 6 minutes longer to deliver than the weekend
- Average cost per order for all orders is $16.50 and the median is $14.14
- Delivery Time may be a factor in the lowest restaurant ratings but not higher ratings
- Order price may have some correlation to restaurant ratings
- Delivery time does not appear to have a relationship to gross revenue
- 58% of orders are repeat customers
- Average number of orders per repeat customer is 2.67
- Total number of repeat customers is 698
- Blue Ribbon Fried Chicken has the highest number of repeat customers at 6
- The most frequent customer placed 13 orders

Multivariate Analysis Continued

Question 17: ***Recommendations***

The data provides a very broad snapshot of information related to FoodHub's business.
We are able to determine what cuisine is most popular, what restaurant receives the most orders, the mean and average delivery times for both weekend and weekdays as well as average revenue/cost per order and cuisine type. Though this data is generally helpful and informative in a broad sense, it does not provide the level of insight and detail that would provide optimal visualization to properly assess the true potential and "what's so" regarding the business delivery process.

In my opinion this data is insufficient to make meaningful and objective, data based decisions for immediate and ongoing strategic implementation.

My first recommendation is to expand the number and types of data points collected. By collecting a broader range of specific data it will allow FoodHub to make better informed decisions in the future. It will also provide for the recognition of present and future trends to help assimilate a model to predict future business opportunities as well as hazards.

Multivariate Analysis Continued

Recommended Data Points for Inclusion:

- Date/Time of Order: record the time and day each order is placed
- Address/Zip Code: record address or at least the zip code of each order delivery
- Distance: record distance from restaurant to delivery address
- Batch or Single Delivery: record how many orders are delivered in each delivery instance
- Cover/Order: How many entrees are included in each order delivered to each address
- Business/Residence: Indicate if the delivery was to a business or a residence
- Holiday: Indicate if the order was placed the day before or the day/weekend of a holiday
- Birthday: If possible, collect just the day and birth month of customer
- Order Contents: Record contents of the order = entrée, appetizer, drink, dessert, salad, etc.
- Lunch/Dinner: When possible distinguish if the order was placed as a lunch or dinner
- Entrée Type: Record if entrée was steak, chicken, hamburger, pizza, sandwich, soup, etc.
- Delivery Time: Prompt & record the specific delivery time estimation provided to customer
- Rating Badge: Promotion system counting number of ratings for incentives to customers

Multivariate Analysis Continued

Summary of recommendations:

There are a number of advantages to expanding the data collection effort. By determining the delivery distance FoodHub can analyze best routes for orders during peak times and combined with the additional information of address and zip code it's possible to increase delivery efficiency, reduce delivery times and identify areas where there is increased order activity. Combined with the current data FoodHub can identify areas of the city that may be interested in specific cuisine types and mount marketing efforts targeting those regions. When collecting addresses we can also determine if specific areas are ordering from specific restaurants based on proximity and/or cuisine type.

Identifying if the orders are delivered in batch or individually will provide additional insight into efficiency and expectations on delivery times. Combined with a system designed to inform the customer of the delivery time and record the estimation will help improve the customers expectation, allow FoodHub to monitor if they are maintaining the estimated delivery times as well as boost ratings by fulfilling on delivery time expectations.

Recording the date and time of the order can provide for specific insight into what times and days are expected for peak delivery activity. By recording the time and day we can provide better delivery coverage for high volume times and days as well as prepare staff for the expected increase in business. If FoodHub can record which orders are place around a holiday, adjustments can be made to prepare for an increase in business. FoodHub can also target and market to those customers and regions where holidays are observed and optimize both marketing revenue and process efficiency.

Multivariate Analysis Continued

Summary of recommendations continued:

Knowing if a delivery is destined for a business or residence can be very useful for projecting future business and resource utilization. For example FoodHub could promote lunch specials to businesses and offer group discounts. In addition if a customer regularly orders from a business address you can potentially offer catering, group lunch discounts or other incentives for a business office celebration or just to increase order number and frequency.

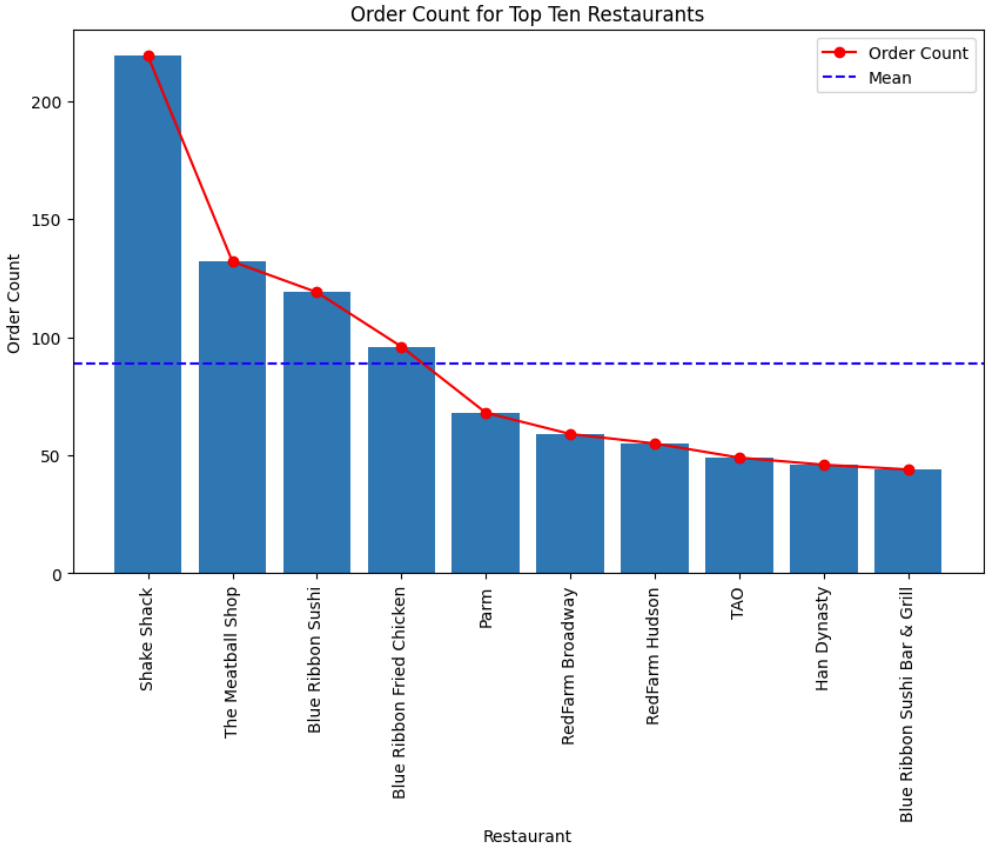
Knowing how many entrees are in a delivery along with the contents of the delivery can provide more insight into production and delivery times as well as potential for suggestive selling and specific offers for future promotions. The more we can understand about each order the more knowledge we have to make better informed decisions in the future. If FoodHub can determine that there is an area of the city that orders 4 entrees of steak dinners, with appetizers, desserts and drinks from a particular restaurant on a regular basis, they have information that can be used to determine future marketing campaigns, advertising expenditures, revenue budgets and necessary resources.

Ratings are a vital part of any delivery program and a greater number of ratings provides for greater feedback and insight into the job the business is doing now as well as how to improve in the future. A 'Ratings Badge' program would provide an incentive for customers to immediately respond to the delivery with a rating. The rating could be a response to a text message or an email. Somewhat similar to a 'rewards card' where the customers acquires points for each rating that will provide for an incentive in the future and creates a sense of urgency and motivation to provide an accurate rating. It would be helpful to gather additional information as part of the rating of 1 -5. Possibly include options such as... Did the delivery arrive on time?, Was the food delivered as ordered? Did the quality of the food meet your expectations? Etc...

APPENDIX

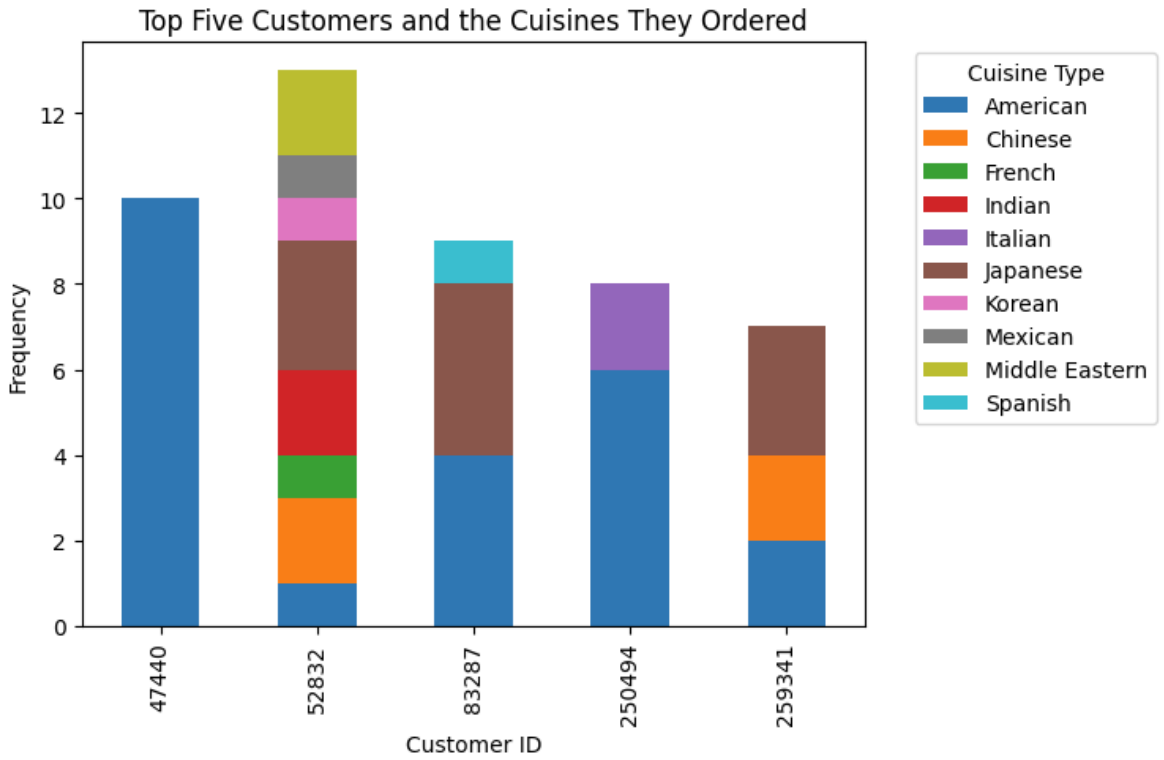
APPENDIX – I

Order Count of Top Ten Grossing Restaurants



APPENDIX – II

Top Five Customers Favorite Cuisine



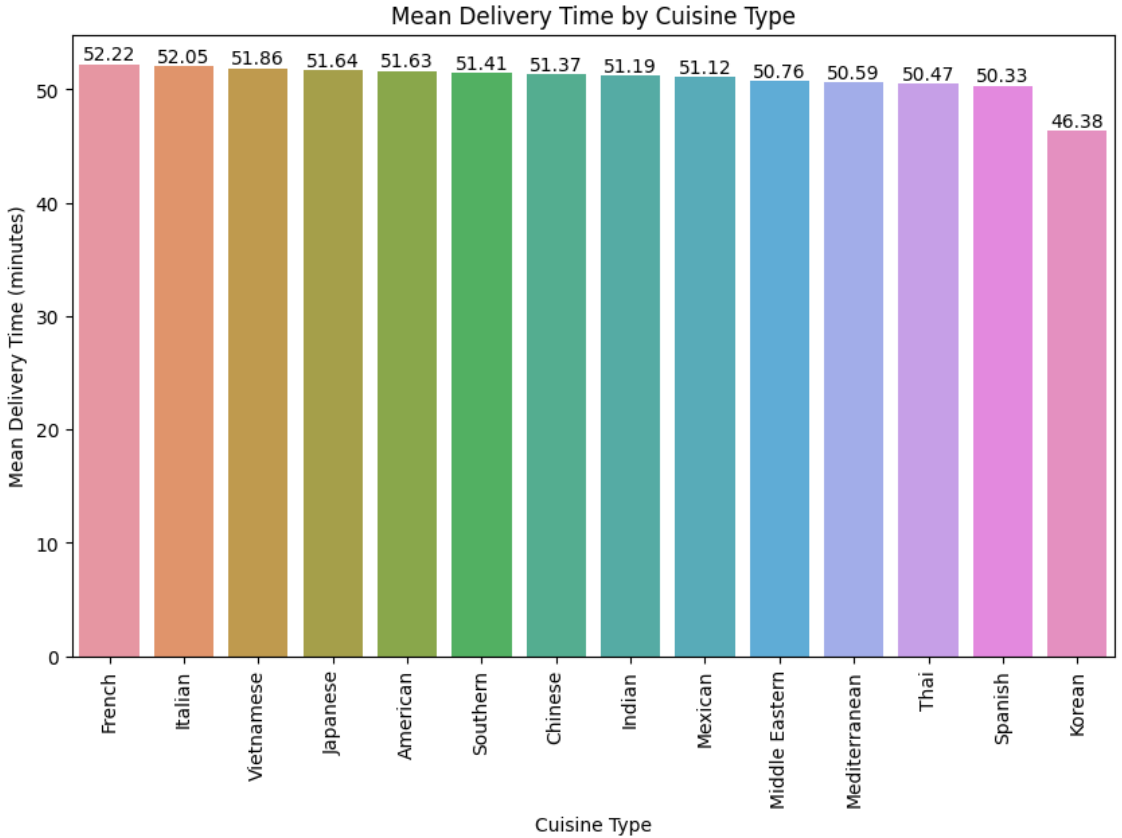
APPENDIX – III

Median Delivery Time For All Orders

APPENDIX – IV

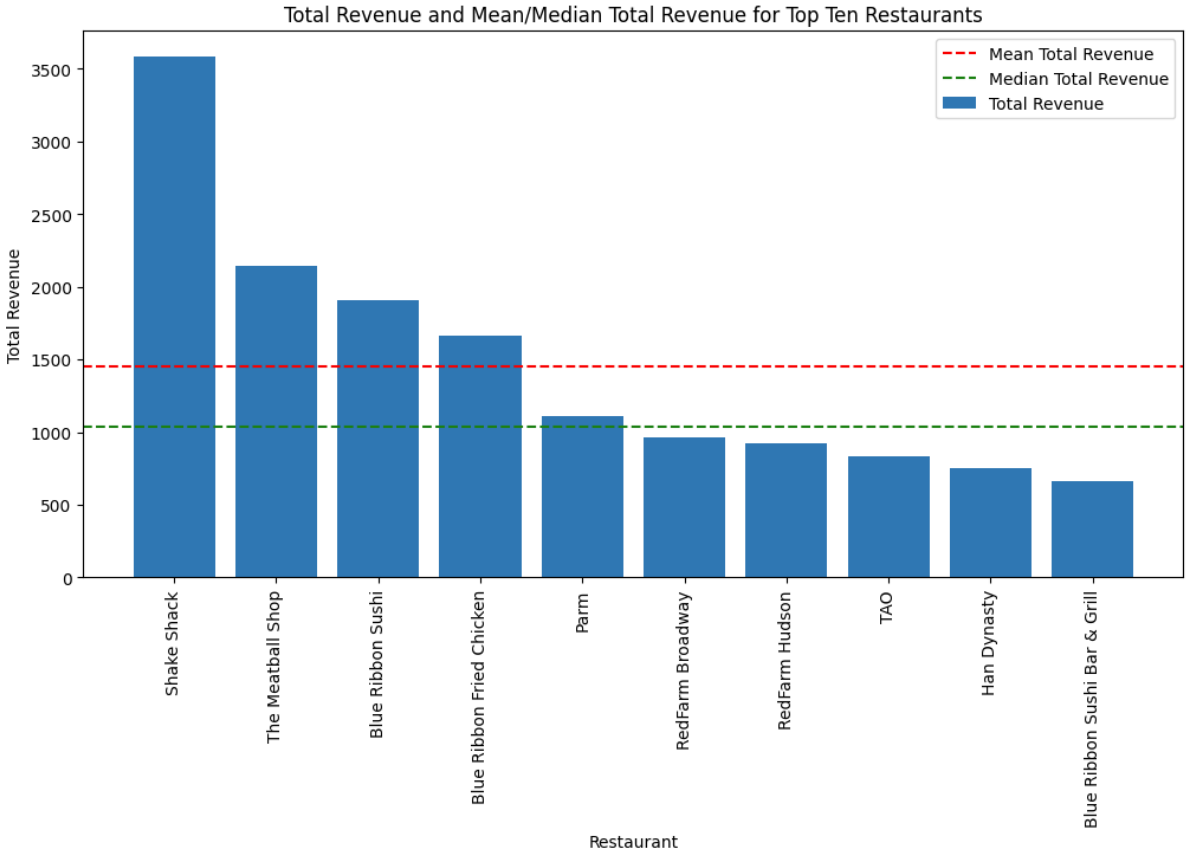
Average Delivery Time By Cuisine

Korean has the fastest average delivery time of all cuisines at 46 minutes



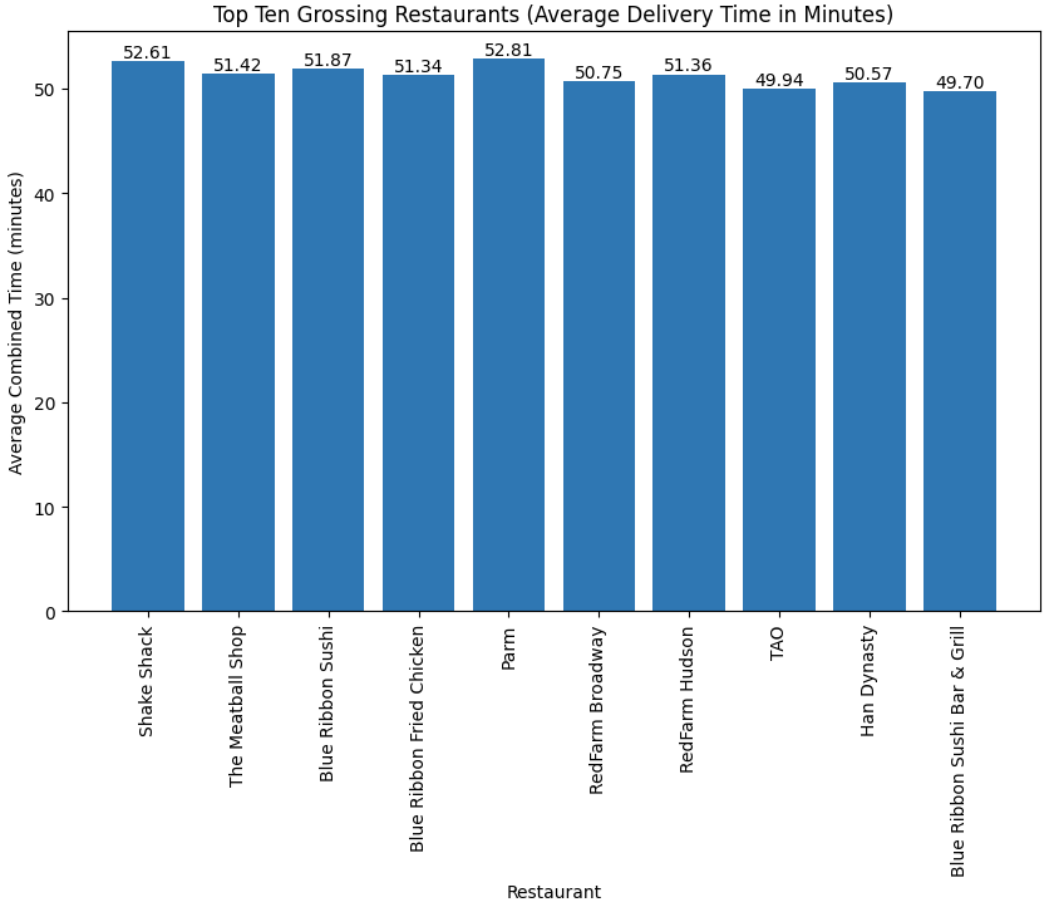
APPENDIX – V

Total Revenue with Mean & Median of Top Ten Grossing Restaurants



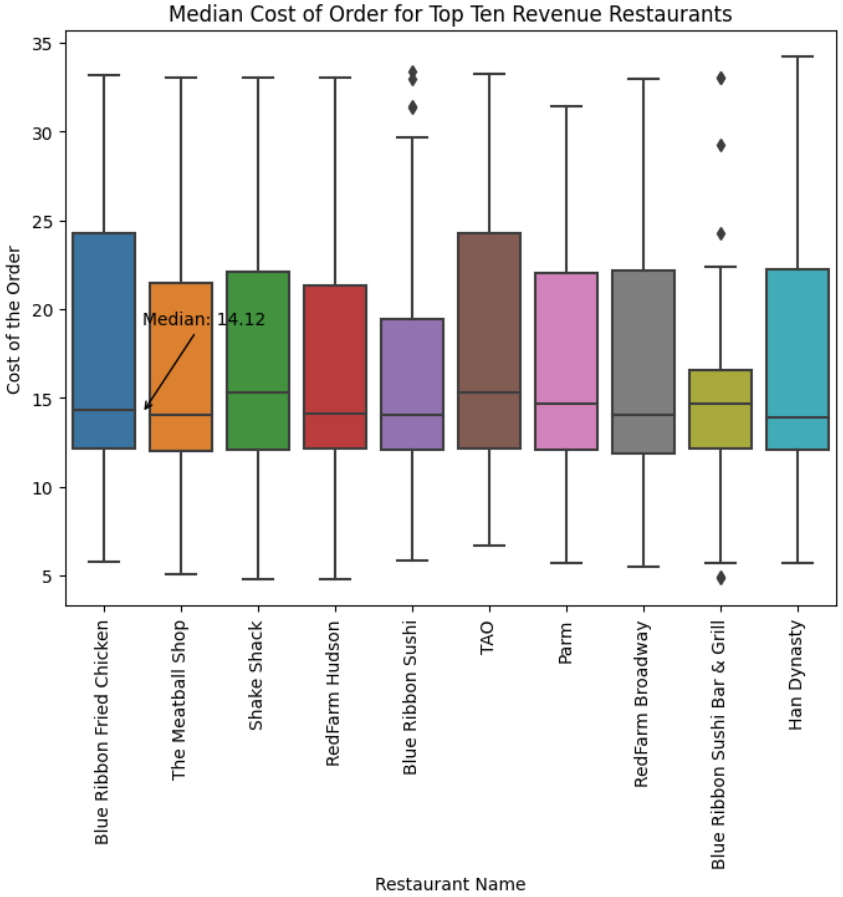
APPENDIX - VI

Average Delivery Time (in minutes) of Top Ten Grossing Restaurants

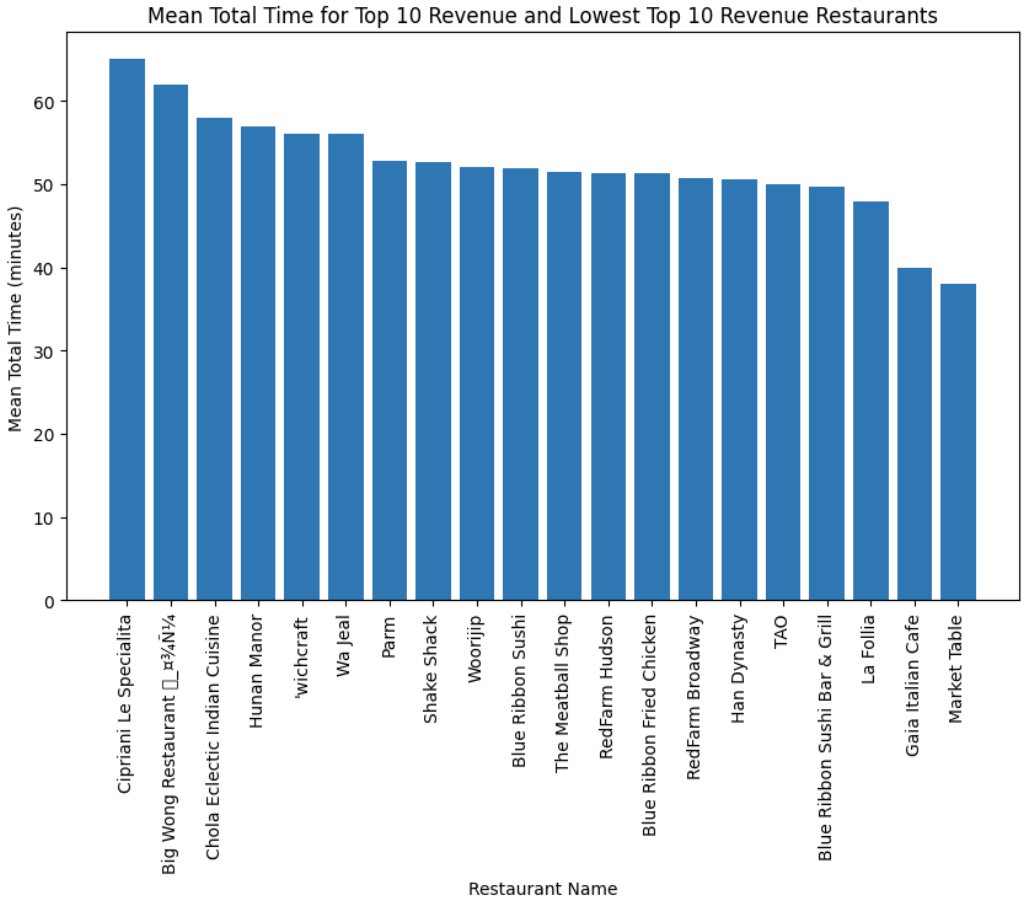


APPENDIX – VII

Median Cost Per Order of Top Ten Grossing Restaurants



APPENDIX – VIII Mean Delivery Times for Ten Highest & Lowest Grossing Restaurants



APPENDIX – IX

Percentage of Total Revenue By Cuisine