Industry: Online Food Ordering and Delivery

By: Victor Abril, Jonathan Diaz, Robert Bernstein

Objective

- Use data from a food ordering and delivery application to analyze customer satisfaction ratings
- Perform an analysis answering the research questions

The Dataset

- Dataset from Kaggle: NYC Restaurants Data - Food Ordering and Delivery
- The dataset (food_order.csv) contains the different data (Columns) related to a food order:
  - order_id
  - customer_id
  - restaurant_name
  - cuisine_type
  - cost
  - day_of_the_week
  - rating
  - food_preparation_time
  - delivery_time

Research Questions

By analyzing the dataset we are going to find insights about:

- What are the top 10 highly rated restaurants (Restaurant vs Rating)
- How does the price paid by the customer affect the rating they gave? (Cost vs Rating)
- How does order completion time affect the ratings? (Order Completion Time vs Rating)
- Which cuisine types were given the highest/lowest ratings? (Cuisine Type vs Rating)

Data Cleaning/Wrangling

Columns	Data Type
order_id	numerical
customer_id	numerical
restaurant_name	categorical
cuisine_type	categorical
cost_of_the_order	numerical
day_of_the_week	categorical
rating	categorical
food_preparation_time	numerical
delivery_time	numerical

Fixed restaurant names

'Big Wong Restaurant \x8c_¤¾Ñ¼' → 'Big Wong Restaurant'

Changed the data type of rating and removed "Not Given" values

['1', '2 ', '3', '4', '5', 'Not Given'] → [1, 2, 3, 4, 5]

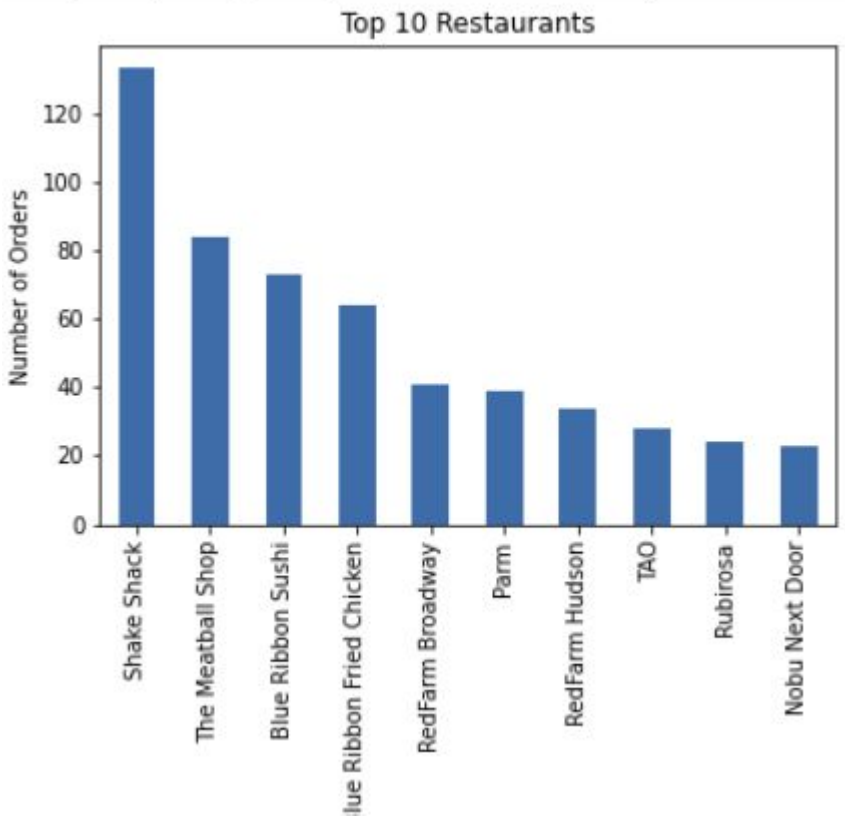
Aggregated Food_preparation_time and delivery_time

Food_preparation_time + delivery_time → order_completion_time

Columns	Data Type
order_id	numerical
customer_id	numerical
restaurant_name	categorical
cuisine_type	categorical
cost_of_the_order	numerical
day_of_the_week	categorical
rating	numerical
order_completion_time	numerical

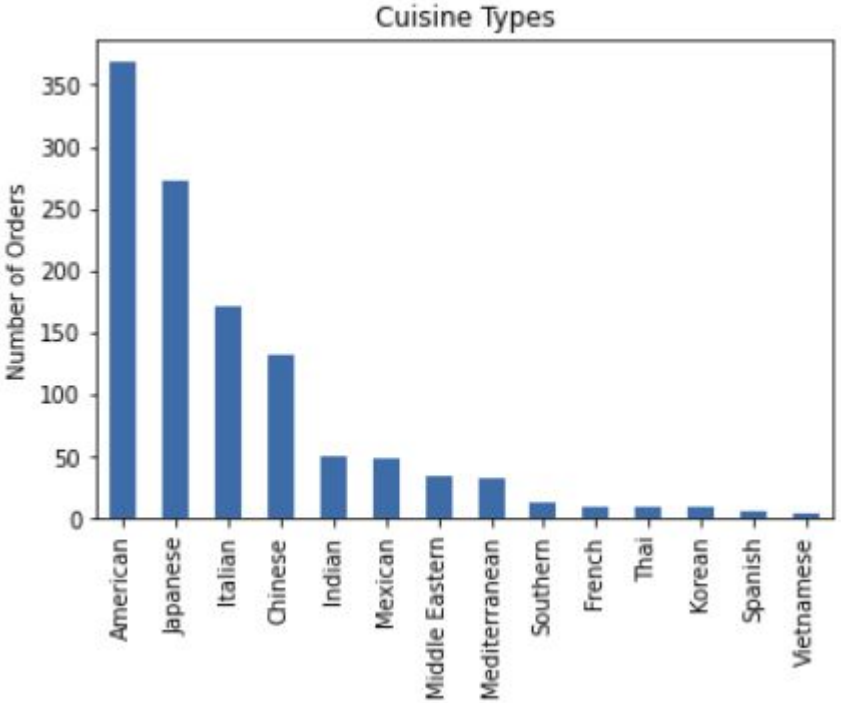
Analysis On The Distribution of The Data

Distribution of restaurant_name



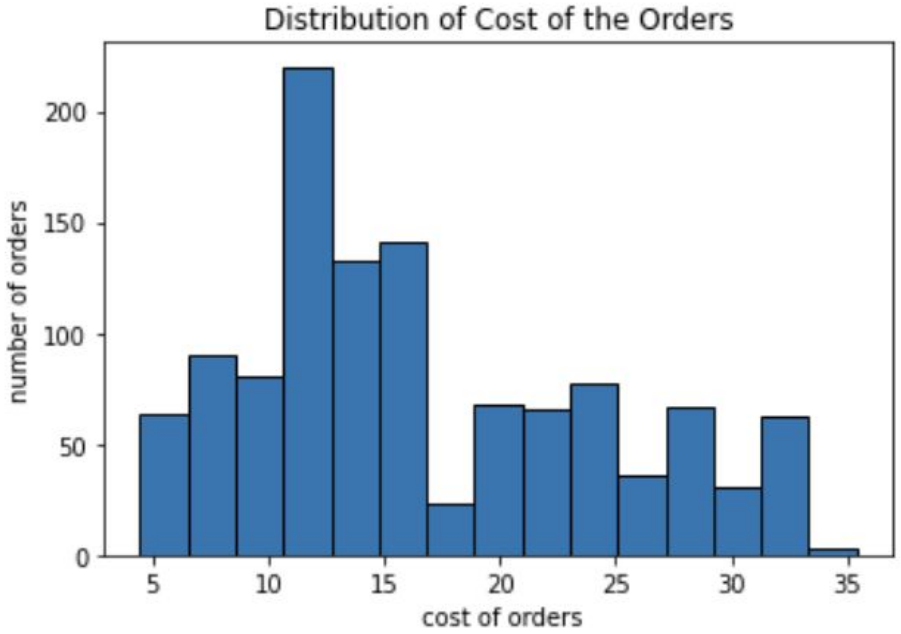
- 156 unique restaurants in the dataset
- Shake Shack has the highest number of orders
- The Meatball Shop comes at 2nd
- The Blue Ribbon Sushi comes at 3rd

Distribution of cuisine_type



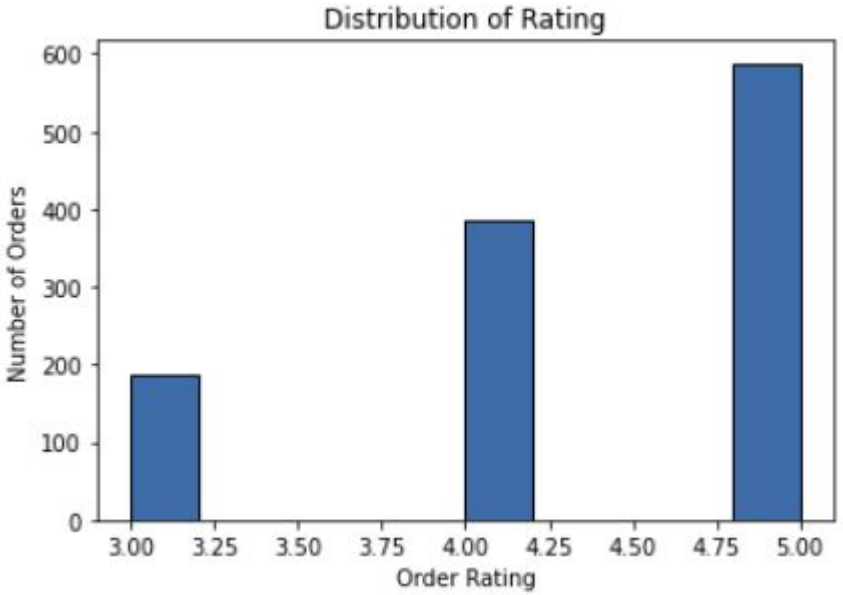
- The top 3 cuisines with the highest number of orders are:
  - American is the most ordered
  - Japanese is the 2nd most ordered
  - Italian is the 3rd most ordered
- The least ordered cuisine is Vietnamese

Distribution of cost_of_the_order



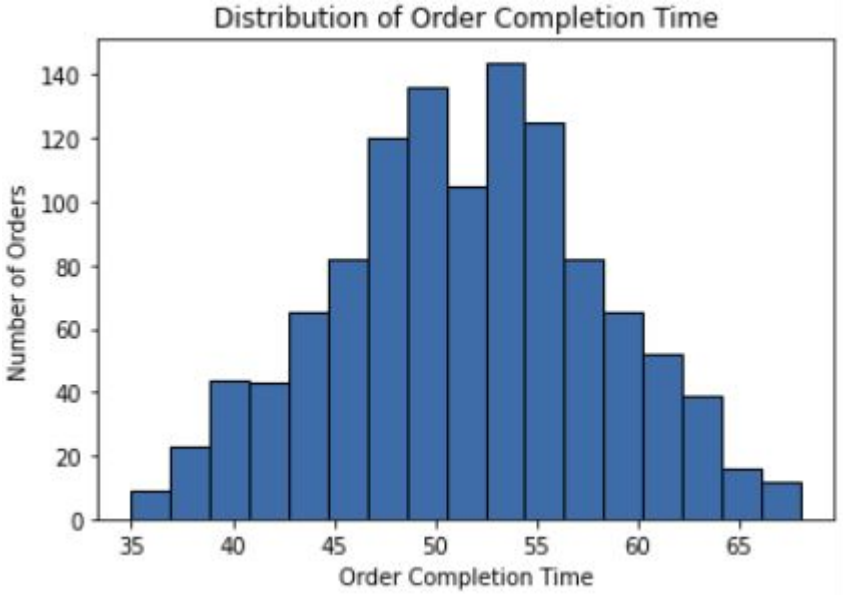
- The histogram indicate skewness towards right.
- The median order cost is approximately 14 dollars, while the mean is at 16.80 dollars.
- The minimum order cost is approximately 4.50 dollars
- The maximum order cost is approximately 35.40 dollars

Distribution of Variable: rating



- More than 75% of the order ratings are between 4 and 5
- The average rating is 4.35
- The minimum rating is 3
- The maximum rating is 5

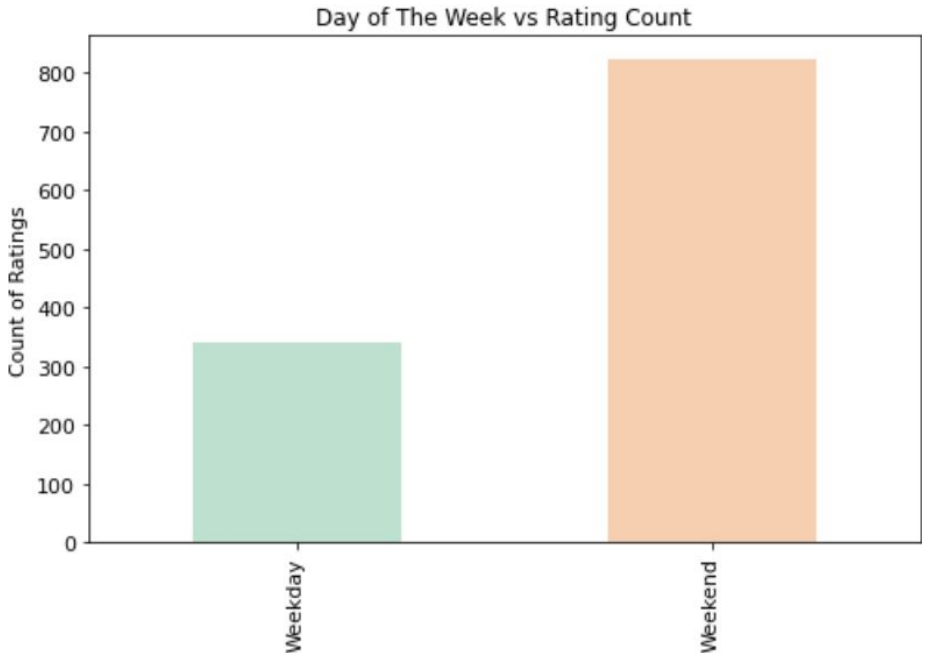
Distribution of order_completion_time



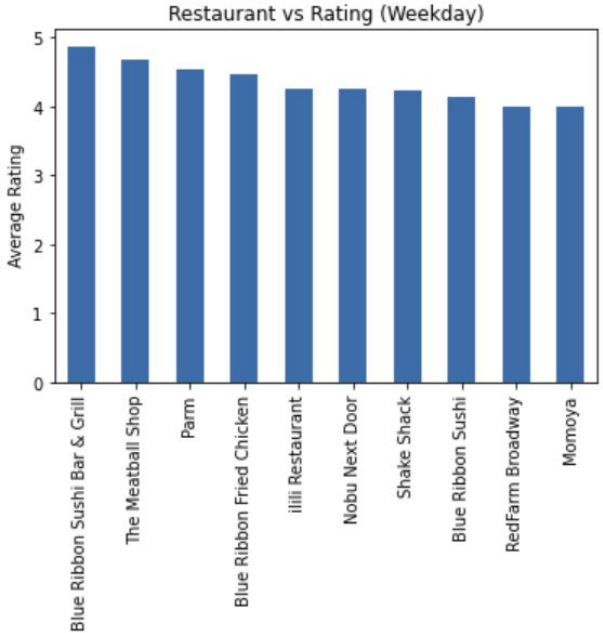
- The histogram indicate a normal distribution
- The median order completion time is approximately 52 minutes , while the mean is at 51.53 minutes.
- The minimum order competition time is approximately 35 minutes
- The maximum order completion time is approximately 68 minutes

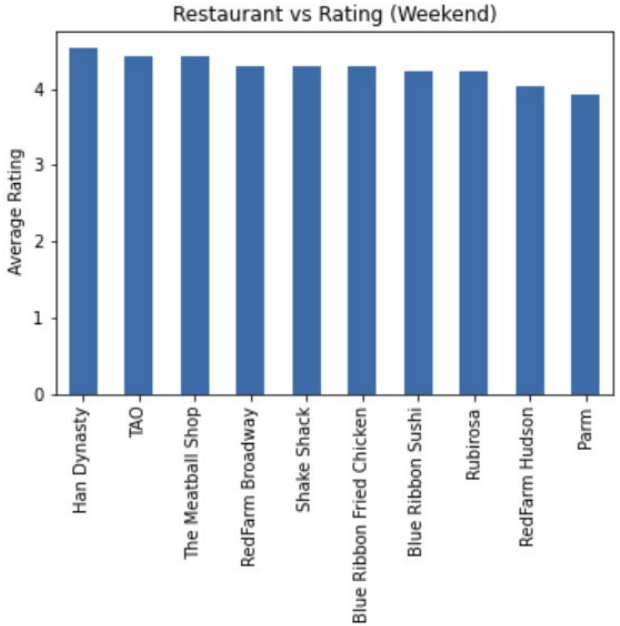
Research Questions

Research Questions

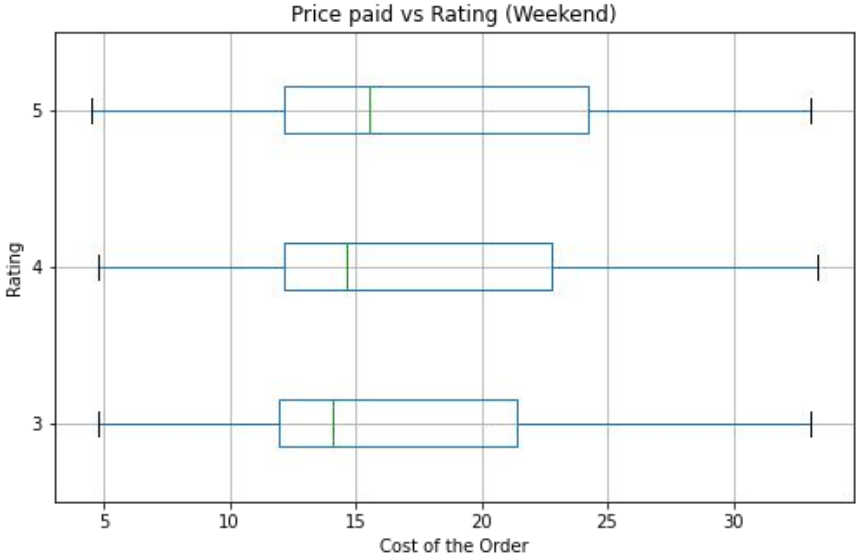


What are the top 10 highly rated Restaurants? (Restaurant vs Rating)

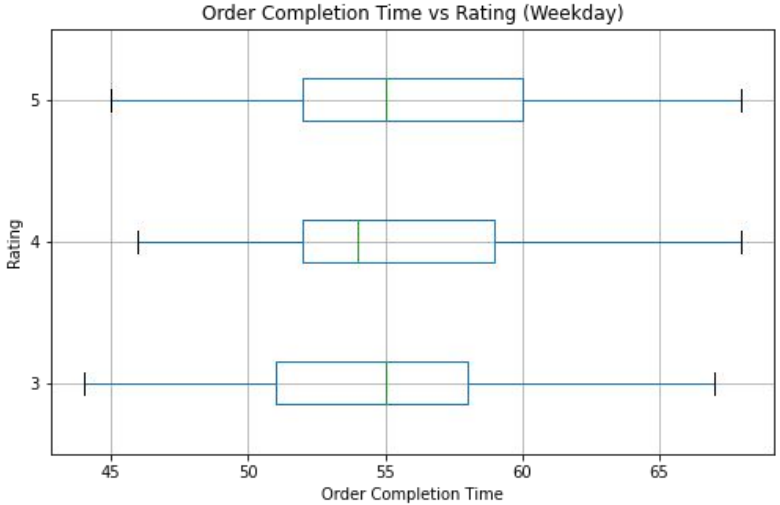


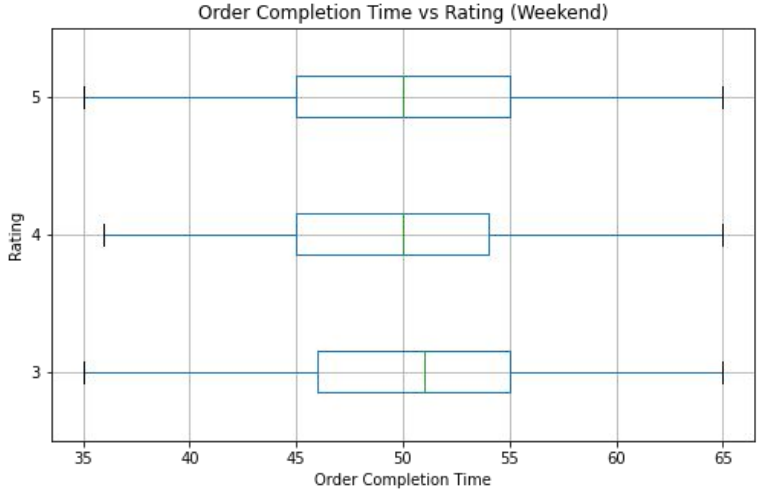


How does the price paid by the customer affect the rating they gave? (Cost vs Rating)

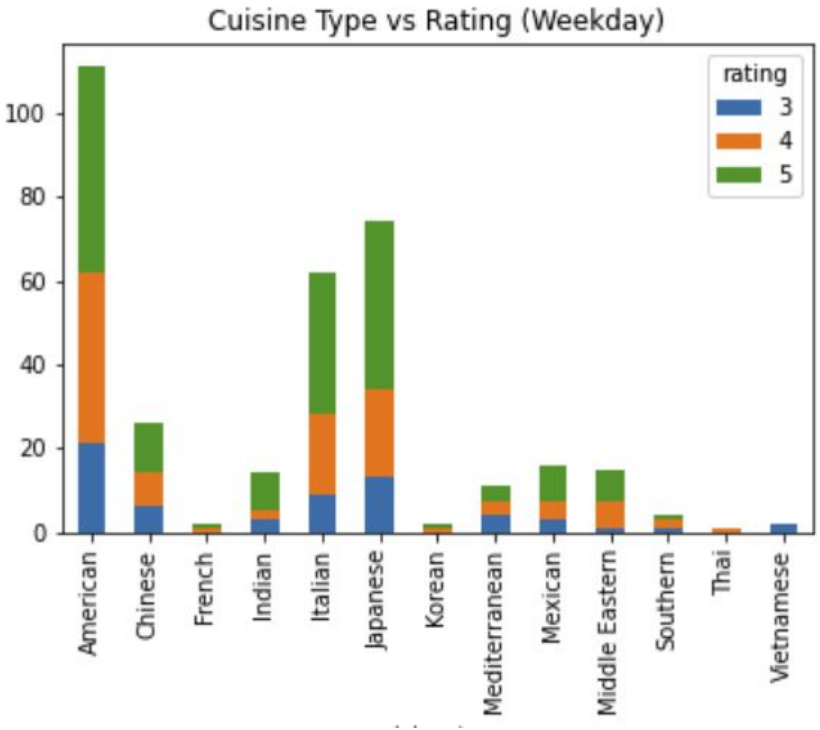


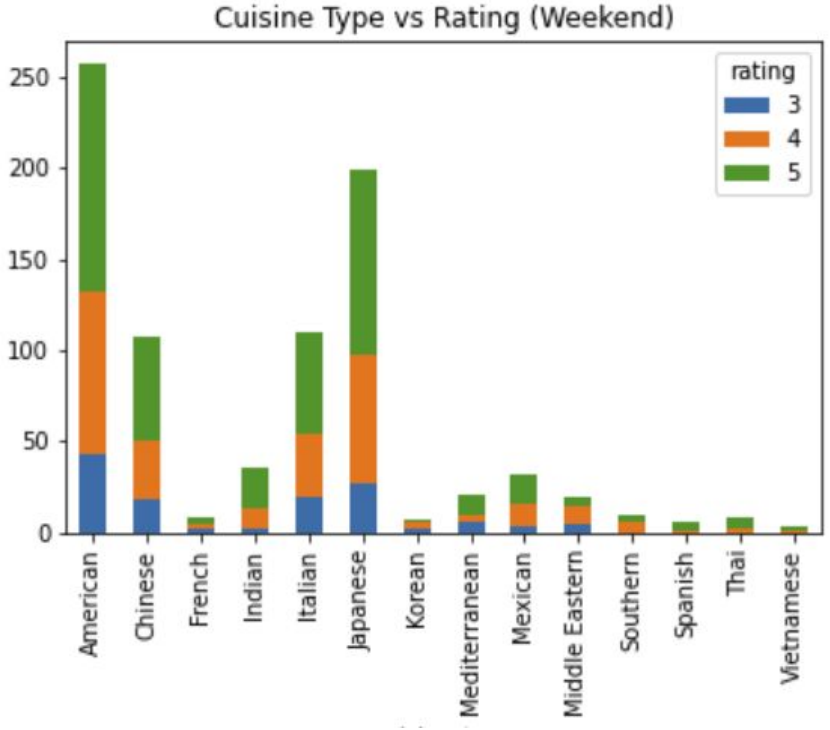
Research Question: How does order completion time affect the ratings? (Order Completion Time vs rating)



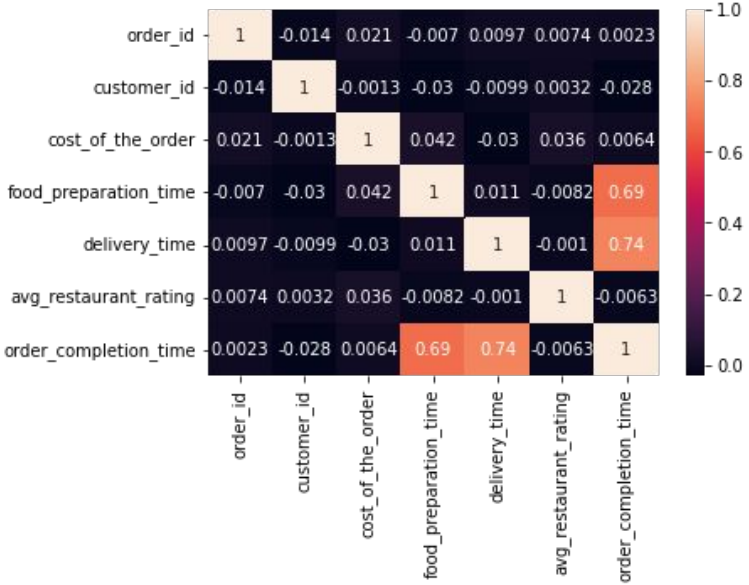


Which cuisine types were given the highest/lowest ratings? (Cuisine Type vs rating)





Correlation matrix



- No significant correlation present

Limitations of the Dataset

- The data does not provide a timestamp specifying the date and time for each order. A more detailed and thorough analysis can be conducted if a timestamp were to be provided
- The data needs to provide a breakdown of the order's cost, allowing us to understand the underlying reason behind the prices for each order
- Approximately 40% of the original data had to be dropped because it did not contain ratings

Key takeaways

- Customers order and leave the most reviews on the weekends
- During the weekends, customers ratings increase as the cost of food increases
- Customers are willing to wait longer during the weekdays than on the weekends based on ratings
- The top rated cuisine types differ between the weekdays and the weekend