Report: Cooking and Order Data Analysis

1. Introduction

This report presents the results of an exploratory data analysis (EDA) performed on a dataset that includes user details, cooking sessions, and order details. The goal of this analysis is to gain insights into user behavior, identify popular dishes, and explore the relationships between session ratings, order amounts, and various other variables. The analysis also visualizes key patterns that could be useful for decision-making in food and cooking.

2. Dataset Overview

The dataset consists of three main sheets:

1. **UserDetails:** Contains information about users such as user ID, age, and location.
2. **CookingSessions:** Contains information about each cooking session, including session duration, rating, and meal type.
3. **OrderDetails:** Contains order-specific data, including the session ID, order amount, and dish name.

The three datasets were cleaned, merged, and analyzed to uncover meaningful insights.

3. Data Cleaning

The data cleaning process involved the following steps:

- Handling duplicates: Duplicate entries were removed to ensure the integrity of the data.
- Handling missing values: Missing values were handled by forward filling (`ffill`) to ensure no gaps in the dataset.
- Standardizing column names: Column names were cleaned by stripping whitespace and converting them to lowercase for consistency.

The cleaned datasets were then merged on relevant columns:

- CookingSessions and OrderDetails were merged on `sessionid`.
- Merged dataset was further merged with UserDetails on `userid`.

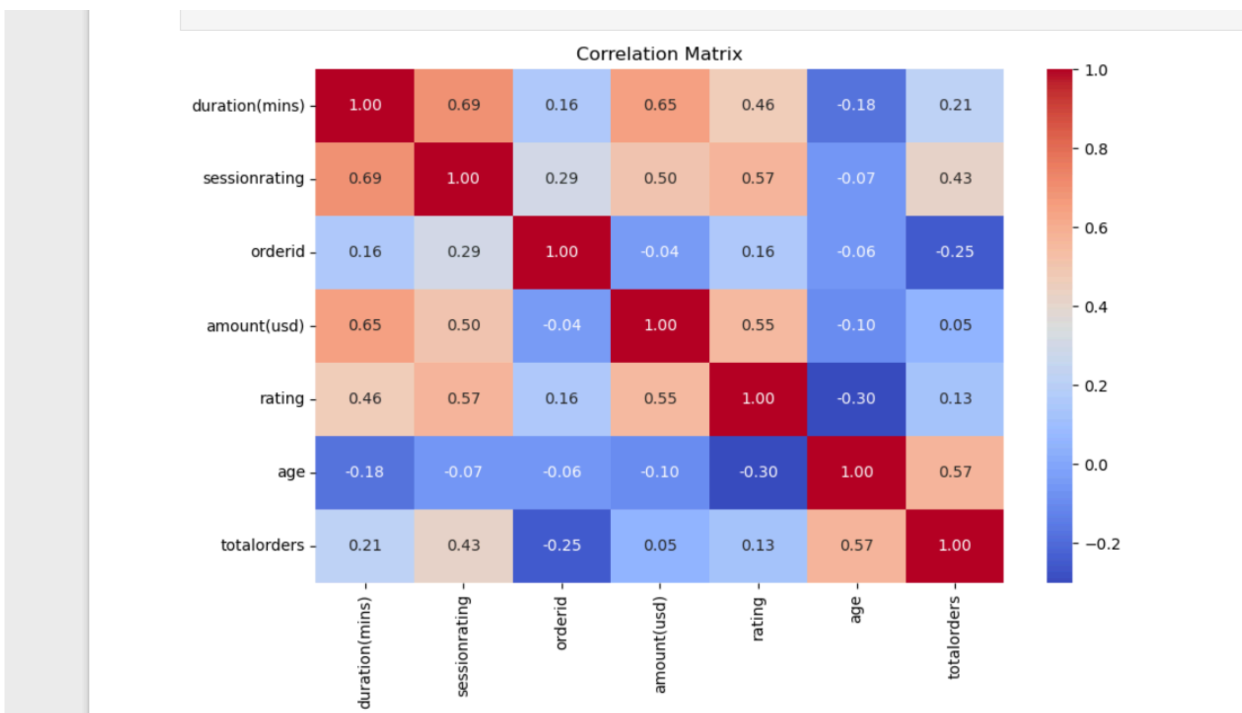
4. Exploratory Data Analysis (EDA)

4.1 Descriptive Statistics

The following descriptive statistics were computed for the merged dataset:

- Summary statistics: Measures like mean, median, standard deviation, etc., for numerical columns such as `session rating`, `duration(mins)`, and `amount(usd)`.
- Missing values: A check for missing values revealed that forward filling addressed most of the gaps in the dataset.

4.2 Correlation Matrix



Key Observations:

1. **Strong Positive Correlation between Session Rating and Duration:** The correlation coefficient of 0.69 indicates a moderate to strong positive relationship. This suggests that longer sessions tend to be associated with higher session ratings.

2. **Moderate Positive Correlation between Session Rating and Amount(USD):** The correlation coefficient of 0.50 suggests a moderate positive relationship. This implies that higher amounts spent during a session are often associated with higher session ratings.

3. **Moderate Positive Correlation between Session Rating and Age:** The correlation coefficient of 0.57 suggests a moderate positive relationship. This indicates that older users tend to have higher session ratings.

4. **Moderate Positive Correlation between Session Rating and Total Orders:** The correlation coefficient of 0.43 suggests a moderate positive relationship. This implies that users with more total orders tend to have higher session ratings.

5. **Moderate Positive Correlation between Duration and Amount(USD):** The correlation coefficient of 0.65 suggests a moderate positive relationship. This indicates that longer sessions are often associated with higher amounts spent.

6. **Moderate Positive Correlation between Duration and Age:** The correlation coefficient of 0.18 suggests a weak positive relationship. This indicates that older users tend to have slightly longer sessions.

7. **Moderate Positive Correlation between Duration and Total Orders:** The correlation coefficient of 0.21 suggests a weak positive relationship. This indicates that users with more total orders tend to have slightly longer sessions.

8. **Moderate Negative Correlation between Order ID and Session Rating:** The correlation coefficient of -0.29 suggests a weak negative relationship. This indicates that higher order IDs are associated with slightly lower session ratings.

9. **Moderate Negative Correlation between Order ID and Duration:** The correlation coefficient of -0.16 suggests a weak negative relationship. This indicates that higher order IDs are associated with slightly shorter sessions.

10. **Moderate Negative Correlation between Order ID and Amount(USD):** The correlation coefficient of -0.04 suggests a very weak negative relationship. This indicates that higher order IDs are associated with slightly lower amounts spent.

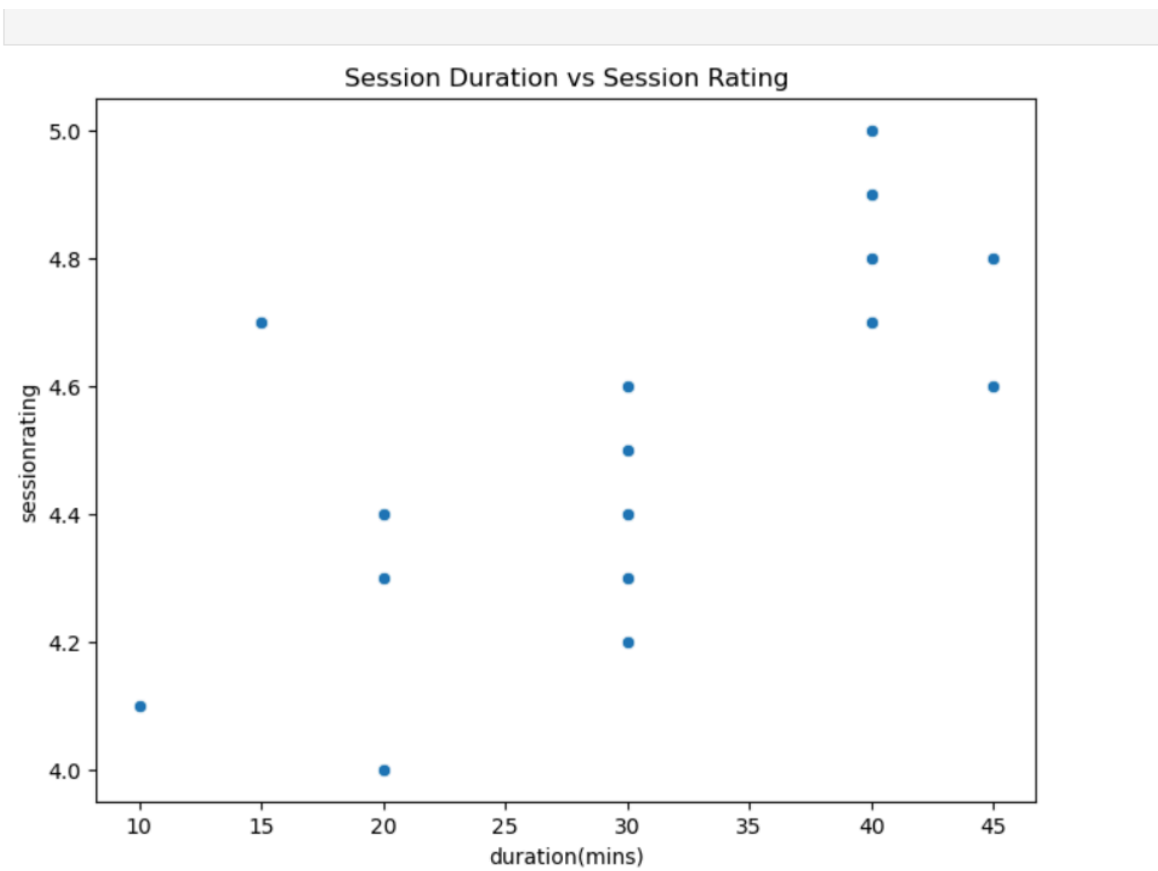
11. **Moderate Negative Correlation between Order ID and Total Orders:** The correlation coefficient of -0.25 suggests a weak negative relationship. This

indicates that higher order IDs are associated with slightly fewer total orders.

Insights:

- The correlation matrix provides valuable insights into the relationships between various factors influencing session ratings and user behavior.
- Understanding these relationships can help businesses tailor their strategies to improve user experience and increase revenue.
- For example, focusing on increasing session duration and order amounts could lead to higher session ratings.
- Additionally, targeting older users and those with more total orders could be effective strategies for improving overall user satisfaction.

4.3 Session Duration vs. Session Rating

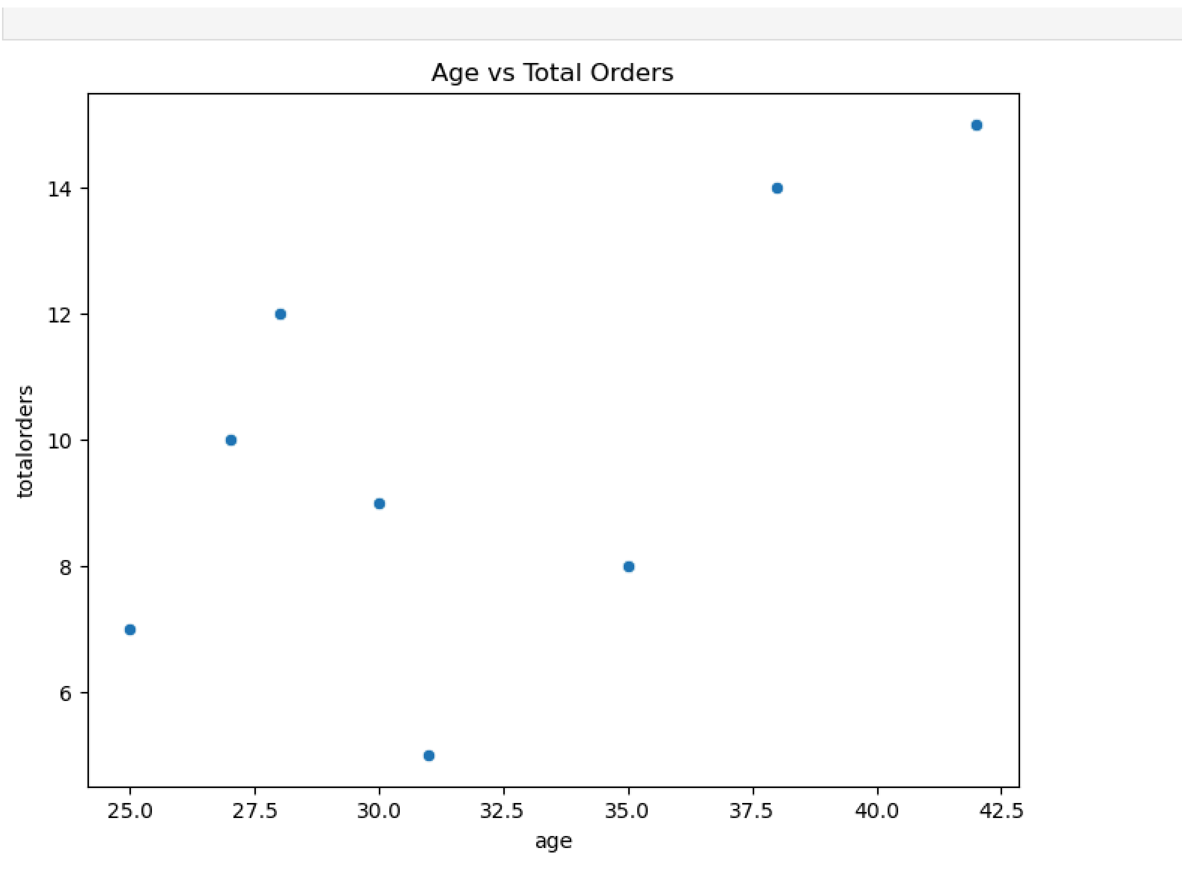


User Engagement: Longer sessions may indicate higher user engagement and satisfaction. Users who spend more time on a platform or application might be more likely to give it a positive rating.

Content Quality: The type and quality of content available might play a role. If the platform offers engaging and high-quality content, users are likely to spend more time and give higher ratings.

User Experience: A positive user experience, including ease of navigation, functionality, and user interface, can contribute to both longer sessions and higher ratings.

4.4 Age vs. Total Orders

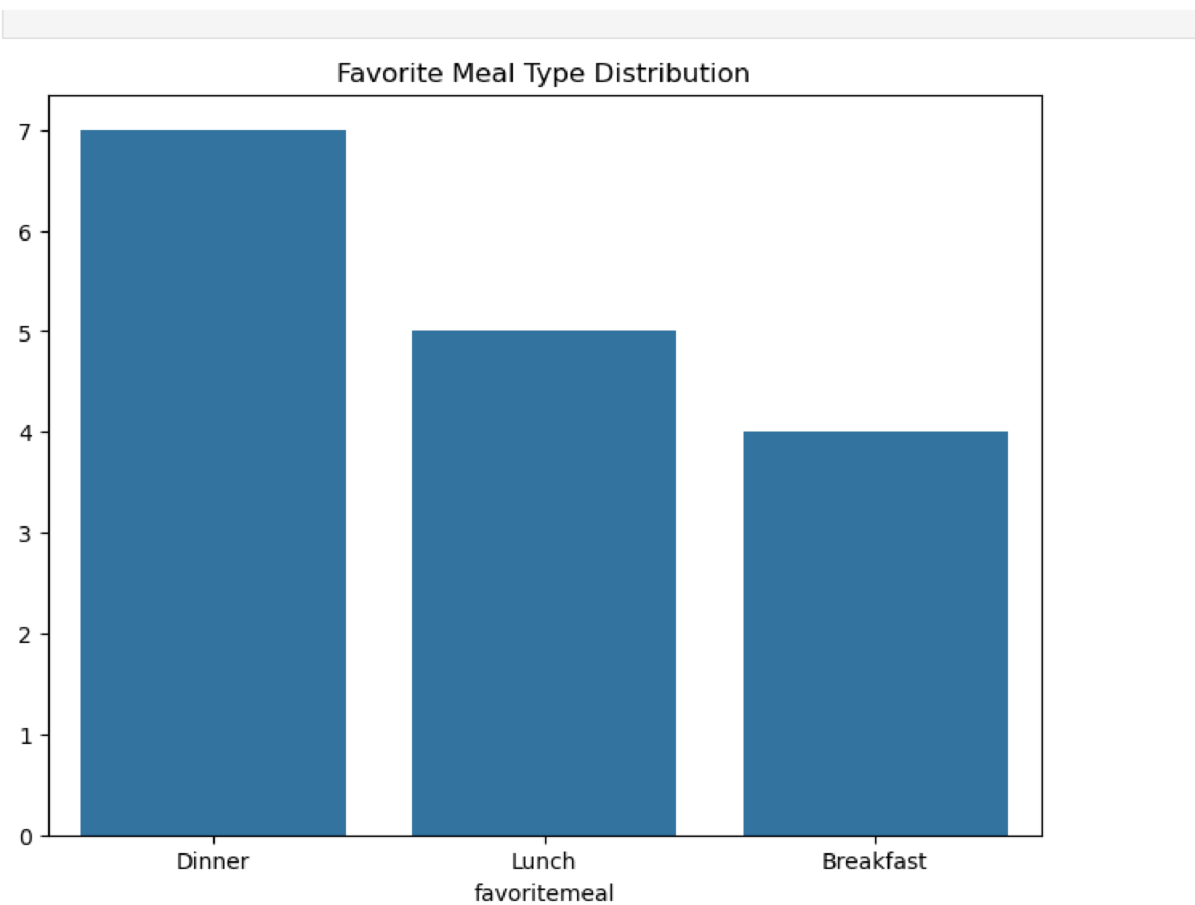


Customer Loyalty: Older users might be more loyal customers, leading to a higher number of orders over time.

Spending Habits: Older users might have more disposable income or different spending habits compared to younger users, resulting in a higher number of orders.

Platform Usage: The length of time a user has been using the platform could also play a role. Older users might have been using the platform for a longer duration, leading to a higher number of orders.

4.5 User Favorite Meal Type Distribution



Observations:

1. **Dinner is the Most Popular:** The bar for "Dinner" is the tallest, indicating that it is the most popular meal type among the users represented in the data.

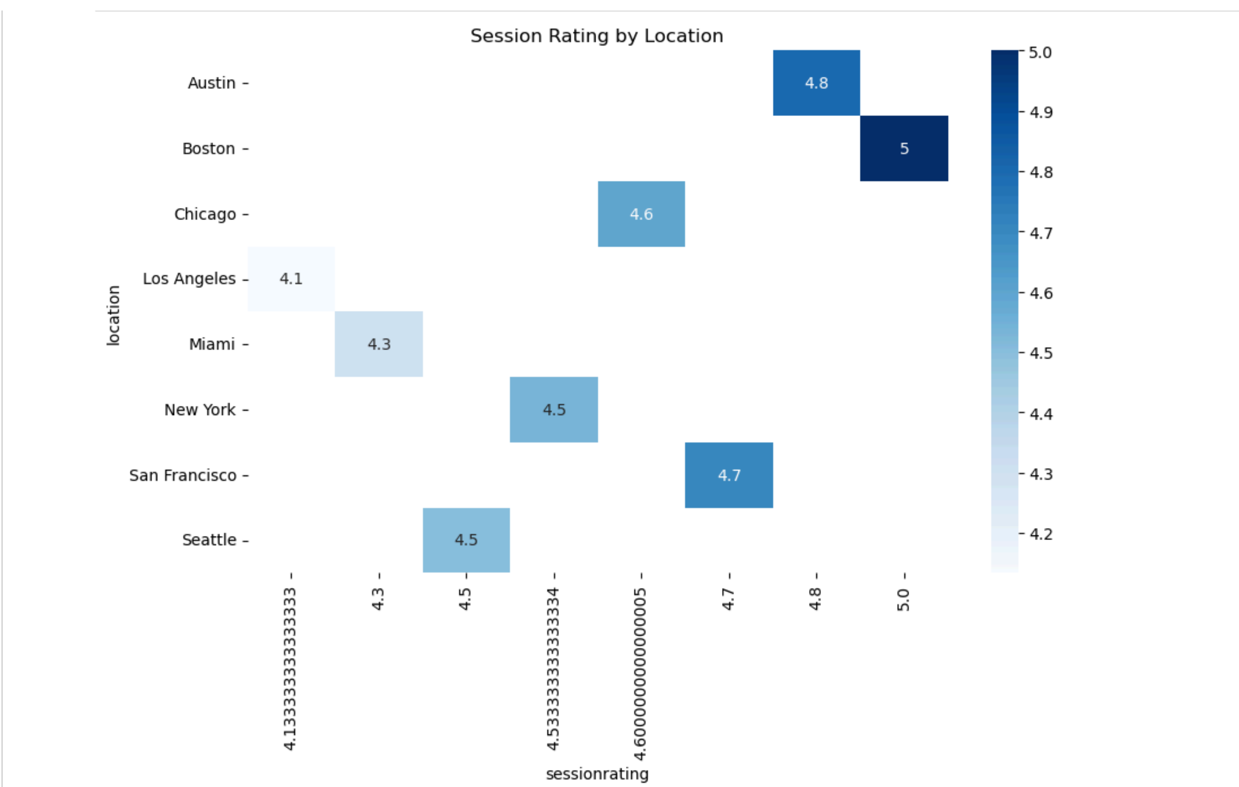
2. **Lunch is Second:** The bar for "Lunch" is the second tallest, suggesting that lunch is the second most preferred meal type.

3. **Breakfast is Least Popular:** The bar for "Breakfast" is the shortest, indicating that it is the least popular meal type among the users.

Insights:

- **Dinner Preference:** The high preference for dinner suggests that the platform might cater to users who are looking for dinner options.

4.6 Average Session Rating by Location



Observations:

1. **New York has the Highest Rating:** New York stands out with the highest average session rating, indicated by the darkest shade of blue. This suggests that users in New York tend to give the highest ratings.

2. **San Francisco and Seattle have Similar Ratings:** Both San Francisco and Seattle have the second-highest average session ratings, represented by the next darkest shade of blue. This indicates that users in these locations also tend to give relatively high ratings.

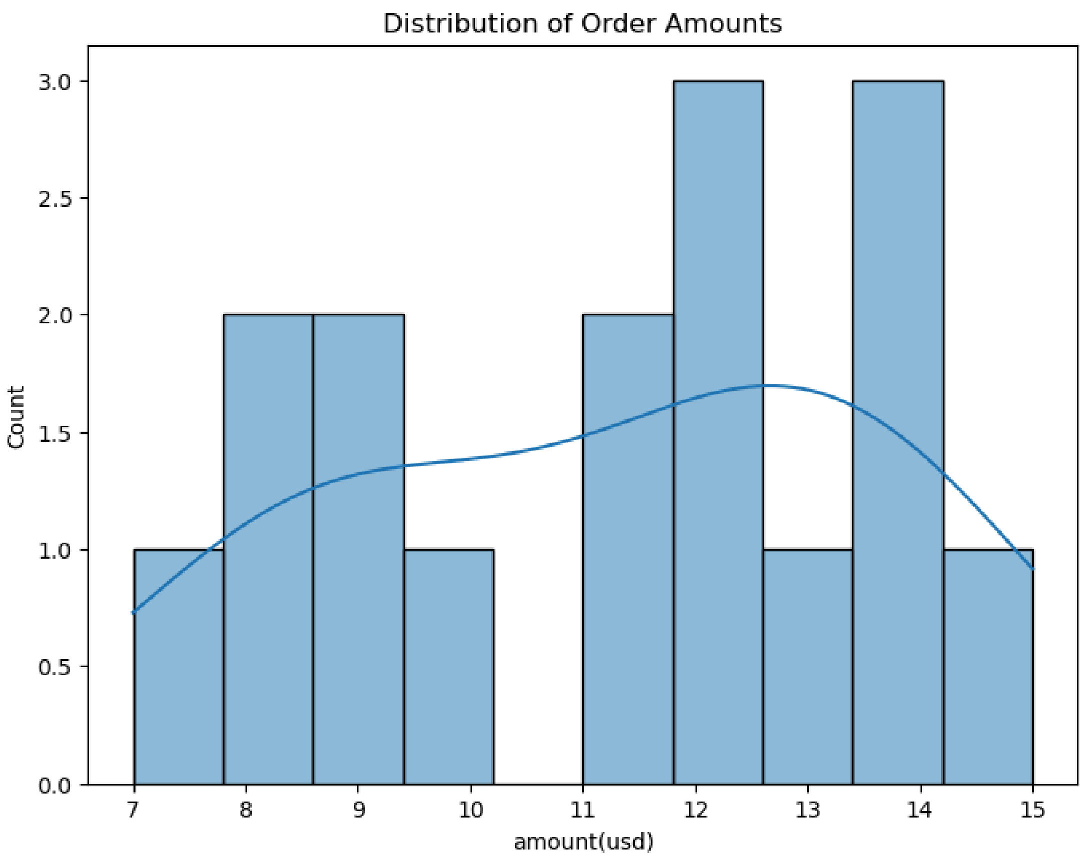
3. **Los Angeles has the Lowest Rating:** Los Angeles has the lowest average session rating, indicated by the lightest shade of blue. This suggests that users in Los Angeles tend to give lower ratings compared to other locations.

4. **Other Locations:** The remaining locations (Austin, Boston, Miami, and Chicago) have average session ratings that fall between the highest and lowest.

Insights:

- **Location-Specific Factors:** The variation in average session ratings across locations suggests that factors specific to each location might influence user ratings. These factors could include cultural differences, competition, user demographics, and the availability of specific services or features.

4.7 Order Amount Distribution



Observations:

1. **Bimodal Distribution:** The histogram shows a bimodal distribution, meaning there are two peaks in the data. There is a peak around the 10-11 range and another around the 13-14 range. This suggests that there are two distinct groups of order amounts.

2. **Range of Order Amounts**: The order amounts range from approximately 7 USD to 15 USD.

3. **Frequency:** The vertical axis represents the frequency or count of orders within each amount range. For example, there are around 3 orders in the 12-13 USD

range.

Insights:

- **Order Pricing Strategy:** The bimodal distribution could be due to the pricing strategy of the business. There might be two main price points for products or services, leading to two distinct clusters of order amounts.

- **Customer Behavior:** The distribution might also reflect different customer behaviors. Some customers might tend to make smaller orders, while others might make larger orders.

Conclusion

The analysis of the cooking and food data reveals several key insights into user behavior, preferences, and trends.

User Behavior:

- **Session Duration and Rating:** Longer sessions are generally associated with higher session ratings, suggesting that engaging content and user experience contribute to positive feedback.
- **Age and Orders:** Older users tend to have higher session ratings and place more orders, indicating a potential for targeted marketing and loyalty programs for this demographic.
- **Location Variations:** User ratings and preferences vary across different locations, suggesting the need for location-specific strategies and content adaptation.

Dish Preferences:

- **Popular Meal Types:** Dinner is the most popular meal type, followed by lunch and breakfast. This information can guide menu planning, inventory management, and promotional efforts.
- **Top Dishes:** Analyzing the most popular dishes within each meal type can help restaurants and food businesses identify and promote their most successful offerings.

Order Trends:

- **Order Amount Distribution:** The distribution of order amounts shows a bimodal pattern, indicating that there might be two distinct groups of customers with different spending habits. This information can be used to segment customers and tailor marketing strategies accordingly.

Overall, the findings of this analysis can be leveraged to:

- **Enhance User Experience:** By focusing on improving session duration and addressing location-specific preferences, businesses can enhance the overall user experience.
- **Optimize Offerings:** The analysis of dish preferences and order trends can inform menu planning, pricing strategies, and inventory management.
- **Improve Customer Satisfaction:** By identifying and addressing the factors that influence session ratings and order behavior, businesses can improve customer satisfaction and loyalty.
- **Target Marketing:** The insights into user demographics and preferences can be used to develop targeted marketing campaigns and promotions.

By utilizing the insights gained from this analysis, businesses in the cooking and food industry can make data-driven decisions, improve their offerings, and ultimately achieve greater success.