```
pip install pandas matplotlib seaborn openpyxl

Requirement already satisfied: pandas in c:\users\suraj_r3os3ti\
appdata\local\programs\python\python312\lib\site-packages (2.2.2)
Requirement already satisfied: matplotlib in c:\users\suraj_r3os3ti\
appdata\local\programs\python\python312\lib\site-packages (3.9.2)
Requirement already satisfied: seaborn in c:\users\suraj_r3os3ti\
appdata\local\programs\python\python312\lib\site-packages (0.13.2)
Requirement already satisfied: openpyxl in c:\users\suraj_r3os3ti\
appdata\local\programs\python\python312\lib\site-packages (3.1.5)
Requirement already satisfied: numpy>=1.26.0 in c:\users\
suraj_r3os3ti\appdata\local\programs\python\python312\lib\site-
packages (from pandas) (1.26.4)
Requirement already satisfied: python-dateutil>=2.8.2 in c:\users\
suraj_r3os3ti\appdata\local\programs\python\python312\lib\site-
packages (from pandas) (2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in c:\users\suraj_r3os3ti\
appdata\local\programs\python\python312\lib\site-packages (from
pandas) (2024.1)
Requirement already satisfied: tzdata>=2022.7 in c:\users\
suraj_r3os3ti\appdata\local\programs\python\python312\lib\site-
packages (from pandas) (2024.1)
Requirement already satisfied: contourpy>=1.0.1 in c:\users\
suraj_r3os3ti\appdata\local\programs\python\python312\lib\site-
packages (from matplotlib) (1.1.1)
Requirement already satisfied: cycler>=0.10 in c:\users\suraj_r3os3ti\
appdata\local\programs\python\python312\lib\site-packages (from
matplotlib) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in c:\users\
suraj_r3os3ti\appdata\local\programs\python\python312\lib\site-
packages (from matplotlib) (4.43.1)
Requirement already satisfied: kiwisolver>=1.3.1 in c:\users\
suraj_r3os3ti\appdata\local\programs\python\python312\lib\site-
packages (from matplotlib) (1.4.5)
Requirement already satisfied: packaging>=20.0 in c:\users\
suraj_r3os3ti\appdata\roaming\python\python312\site-packages (from
matplotlib) (23.2)
Requirement already satisfied: pillow>=8 in c:\users\suraj_r3os3ti\
appdata\local\programs\python\python312\lib\site-packages (from
matplotlib) (10.1.0)
Requirement already satisfied: pyparsing>=2.3.1 in c:\users\
suraj_r3os3ti\appdata\local\programs\python\python312\lib\site-
packages (from matplotlib) (3.1.1)
Requirement already satisfied: et-xmlfile in c:\users\suraj_r3os3ti\
appdata\local\programs\python\python312\lib\site-packages (from
openpyxl) (1.1.0)
Requirement already satisfied: six>=1.5 in c:\users\suraj_r3os3ti\
appdata\local\programs\python\python312\lib\site-packages (from
python-dateutil>=2.8.2->pandas) (1.16.0)
Note: you may need to restart the kernel to use updated packages.
```

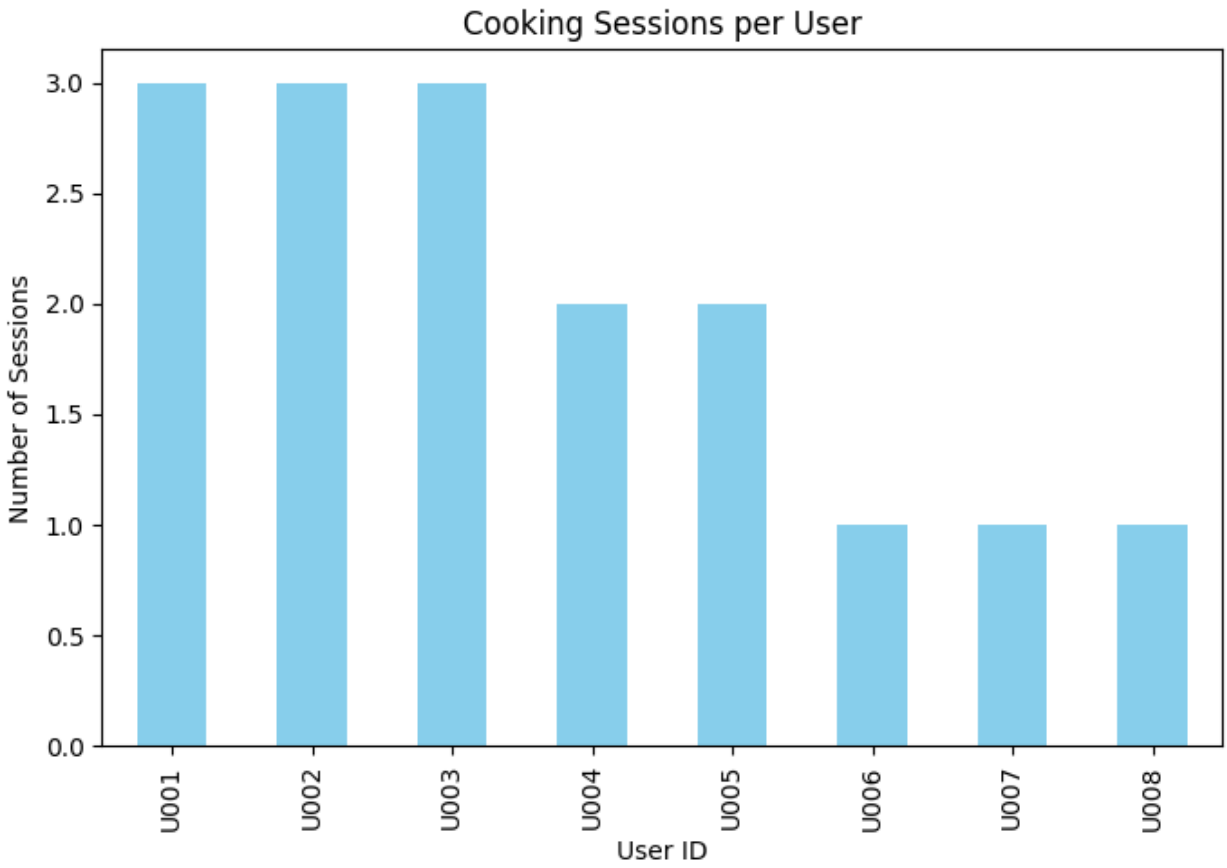
```
WARNING: Ignoring invalid distribution ~ (c:\Users\suraj_r3os3ti\
AppData\Local\Programs\Python\Python312\Lib\site-packages)
WARNING: Ignoring invalid distribution ~ (c:\Users\suraj_r3os3ti\
AppData\Local\Programs\Python\Python312\Lib\site-packages)
WARNING: Ignoring invalid distribution ~ (c:\Users\suraj_r3os3ti\
AppData\Local\Programs\Python\Python312\Lib\site-packages)
```

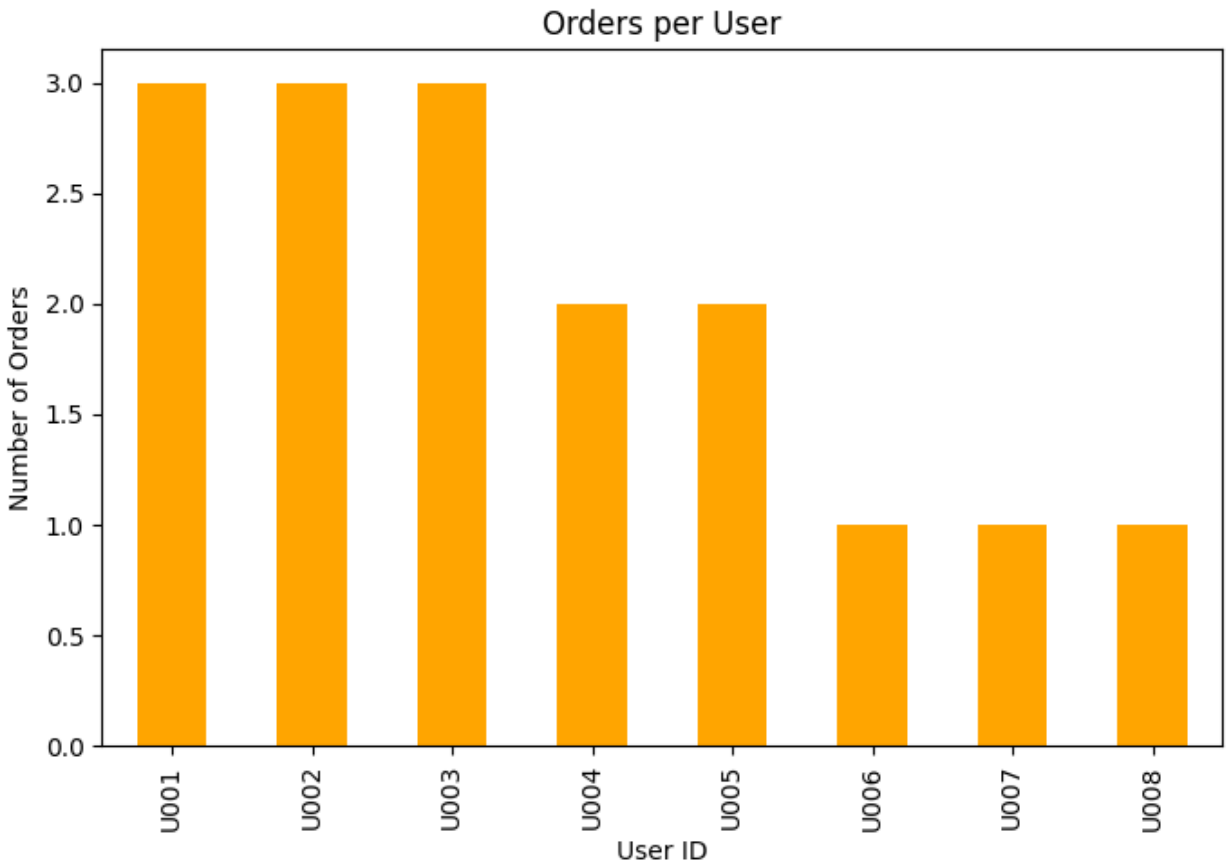
```python
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Load the datasets
user_details = pd.read_excel("Data Analyst Intern Assignment -
Excel.xlsx", sheet_name="UserDetails.csv")
cooking_sessions = pd.read_excel("Data Analyst Intern Assignment -
Excel.xlsx", sheet_name="CookingSessions.csv")
order_details = pd.read_excel("Data Analyst Intern Assignment -
Excel.xlsx", sheet_name="OrderDetails.csv")

# Data Cleaning
def clean_data(df):
    # Remove duplicates
    df.drop_duplicates(inplace=True)
    # Handle missing values
    df.fillna("Unknown", inplace=True)
    return df

user_details = clean_data(user_details)
cooking_sessions = clean_data(cooking_sessions)
order_details = clean_data(order_details)

# Analysis 1: Relationship between cooking sessions and user orders
cooking_trend = cooking_sessions["User ID"].value_counts()
order_trend = order_details["User ID"].value_counts()

# Plot Cooking Sessions
plt.figure(figsize=(8, 5))
cooking_trend.plot(kind="bar", color="skyblue")
plt.title("Cooking Sessions per User")
plt.xlabel("User ID")
plt.ylabel("Number of Sessions")
plt.show()

# Plot User Orders
plt.figure(figsize=(8, 5))
order_trend.plot(kind="bar", color="orange")
plt.title("Orders per User")
plt.xlabel("User ID")
plt.ylabel("Number of Orders")
plt.show()
```

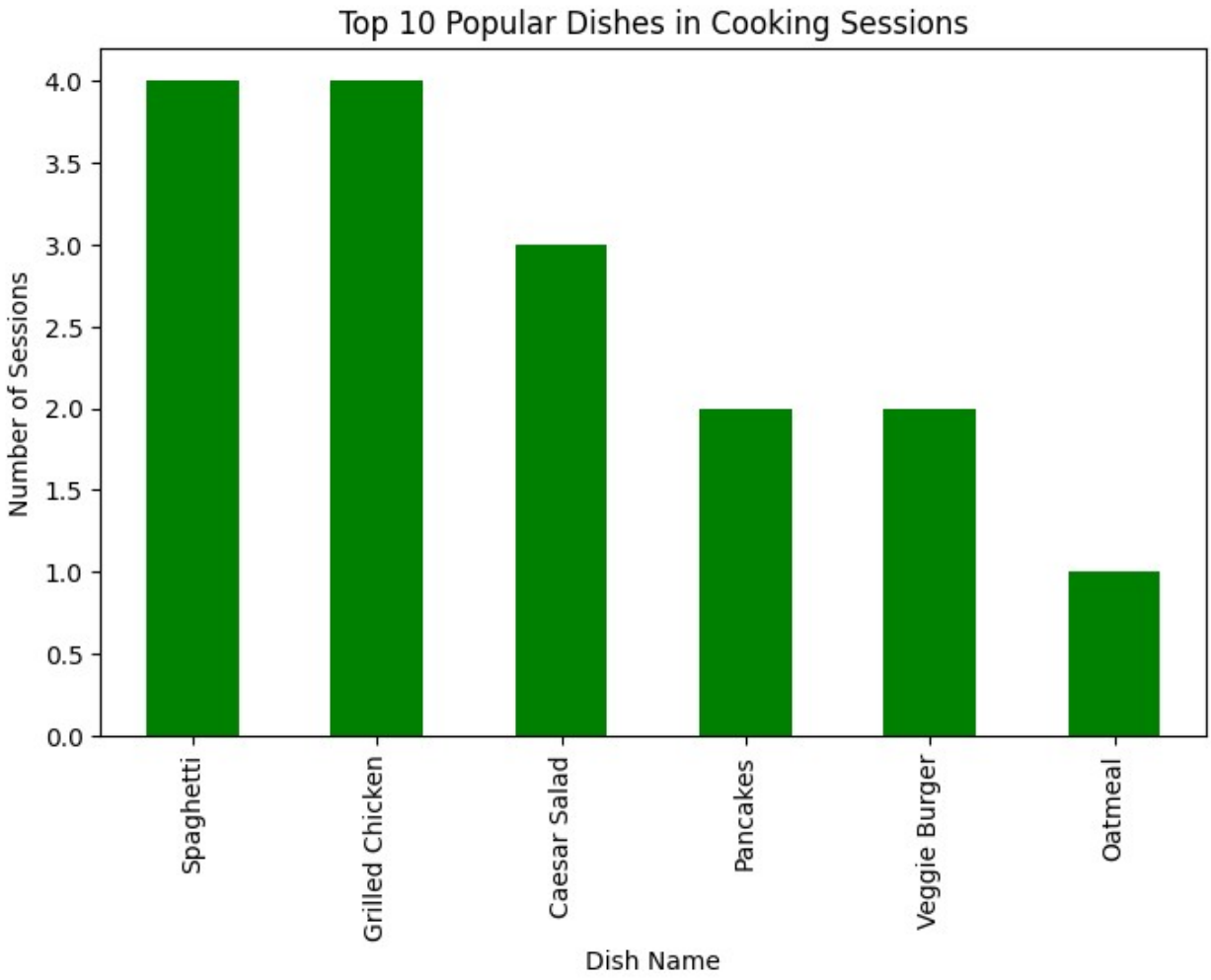
```
# Analysis 2: Popular Dishes
popular_cooking_dishes = cooking_sessions["Dish
Name"].value_counts().head(10)
popular_order_dishes = order_details["Dish
Name"].value_counts().head(10)

# Plot Popular Dishes in Cooking Sessions
plt.figure(figsize=(8, 5))
popular_cooking_dishes.plot(kind="bar", color="green")
plt.title("Top 10 Popular Dishes in Cooking Sessions")
plt.xlabel("Dish Name")
plt.ylabel("Number of Sessions")
plt.show()

# Plot Popular Dishes in Orders
plt.figure(figsize=(8, 5))
popular_order_dishes.plot(kind="bar", color="purple")
plt.title("Top 10 Popular Dishes in Orders")
plt.xlabel("Dish Name")
plt.ylabel("Number of Orders")
plt.show()

# Analysis 3: Demographic factors influencing user behavior
sns.boxplot(x="Favorite Meal", y="Age", data=user_details)
plt.title("Favorite Meal by Age Group")
plt.show()

# Analysis 4: Session ratings
sns.histplot(cooking_sessions["Session Rating"], kde=True,
color="blue", bins=10)
plt.title("Distribution of Session Ratings")
plt.xlabel("Session Rating")
plt.ylabel("Frequency")
plt.show()

# Revenue Analysis
order_details["Order Date"] = pd.to_datetime(order_details["Order
Date"])
revenue_trend = order_details.groupby("Order Date")["Amount
(USD)"].sum()

# Plot Revenue Trends
plt.figure(figsize=(10, 5))
revenue_trend.plot(kind="line", color="red")
plt.title("Daily Revenue Trends")
plt.xlabel("Date")
plt.ylabel("Revenue (USD)")
plt.show()
```

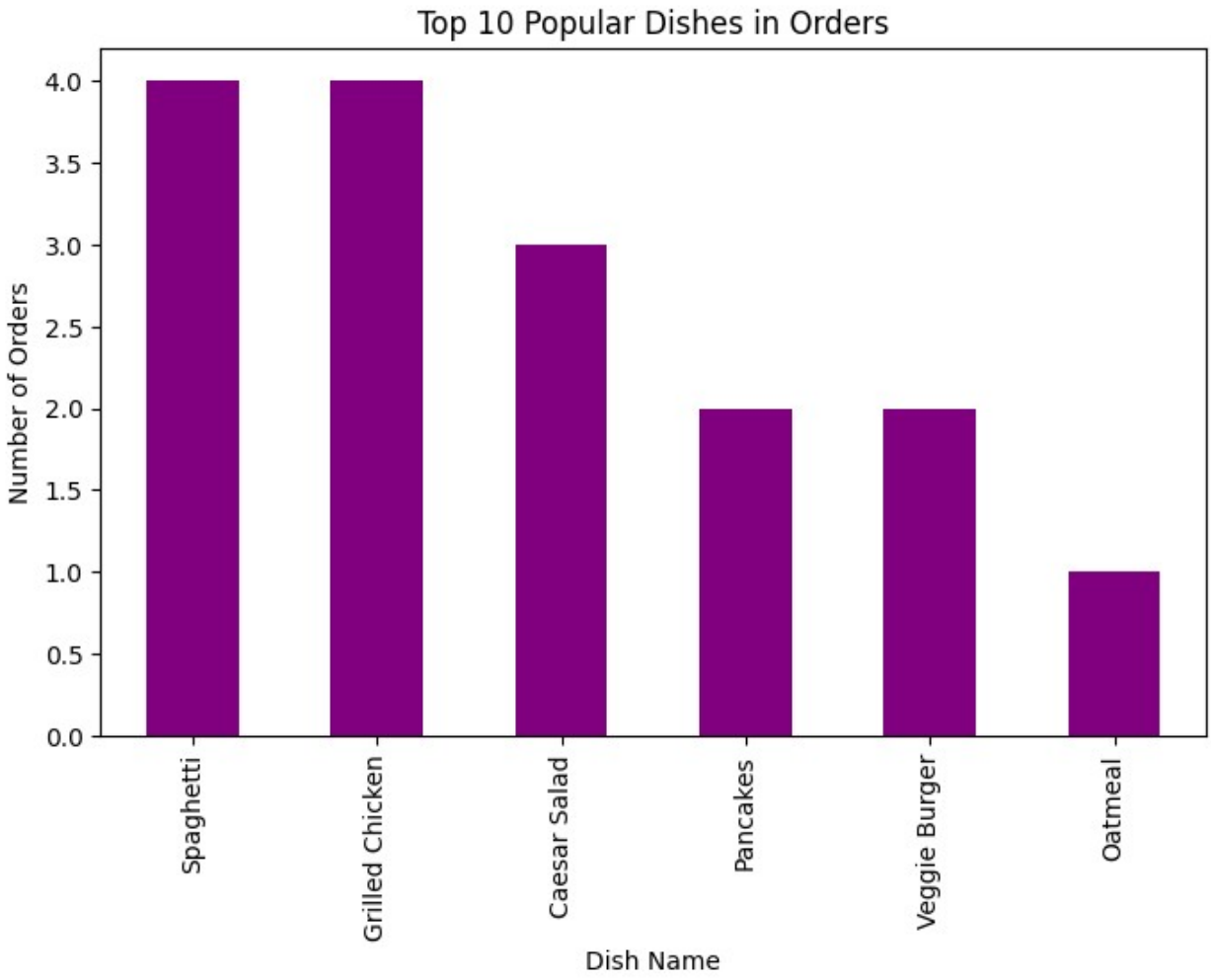
```python
# Writing a Summary Report
summary = {
    "Top Cooking Dishes": popular_cooking_dishes.to_dict(),
    "Top Ordered Dishes": popular_order_dishes.to_dict(),
    "Top Revenue Day": revenue_trend.idxmax(),
    "Maximum Revenue": revenue_trend.max(),
}
with open("Summary_Report.txt", "w") as file:
    for key, value in summary.items():
        file.write(f"{key}: {value}\n")

print("Analysis and visualizations completed. Summary report saved as
'Summary_Report.txt'.")
```

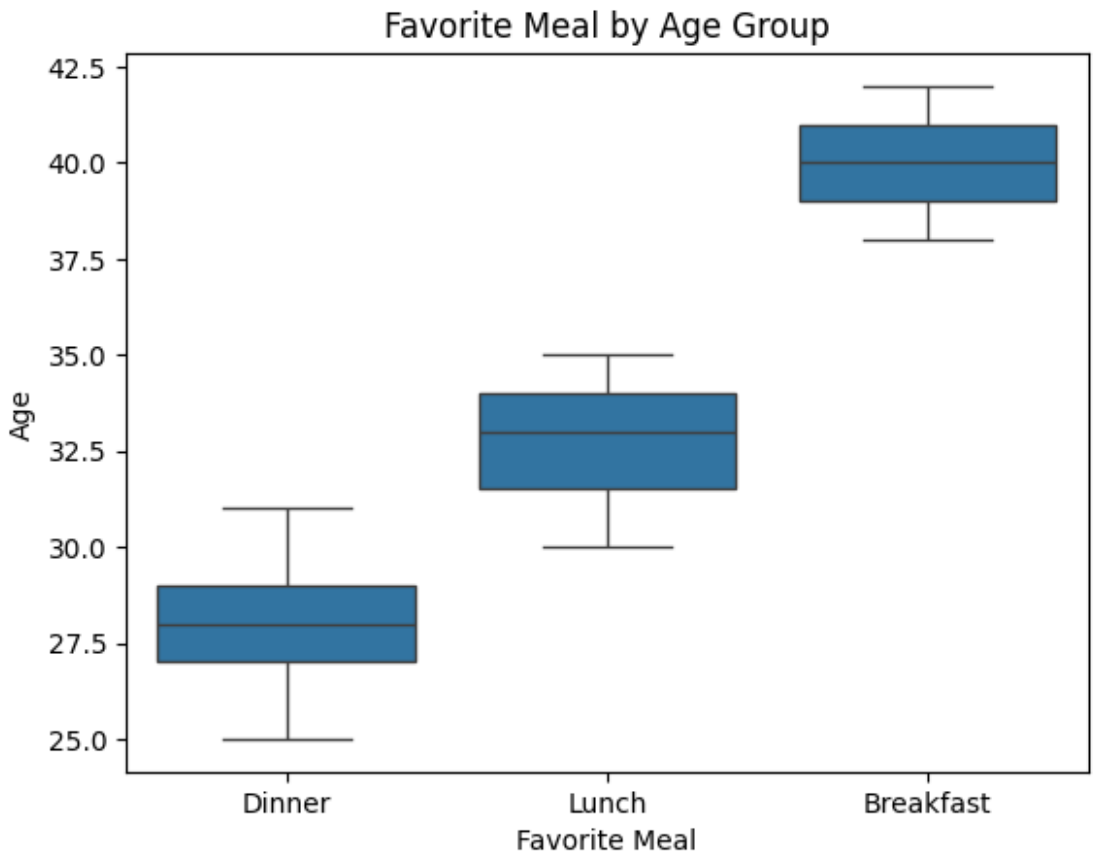
```
C:\Users\suraj_r3os3ti\AppData\Local\Temp\
ipykernel_4544\500228922.py:15: FutureWarning: Setting an item of
incompatible dtype is deprecated and will raise an error in a future
version of pandas. Value 'Unknown' has dtype incompatible with
float64, please explicitly cast to a compatible dtype first.
  df.fillna("Unknown", inplace=True)
```

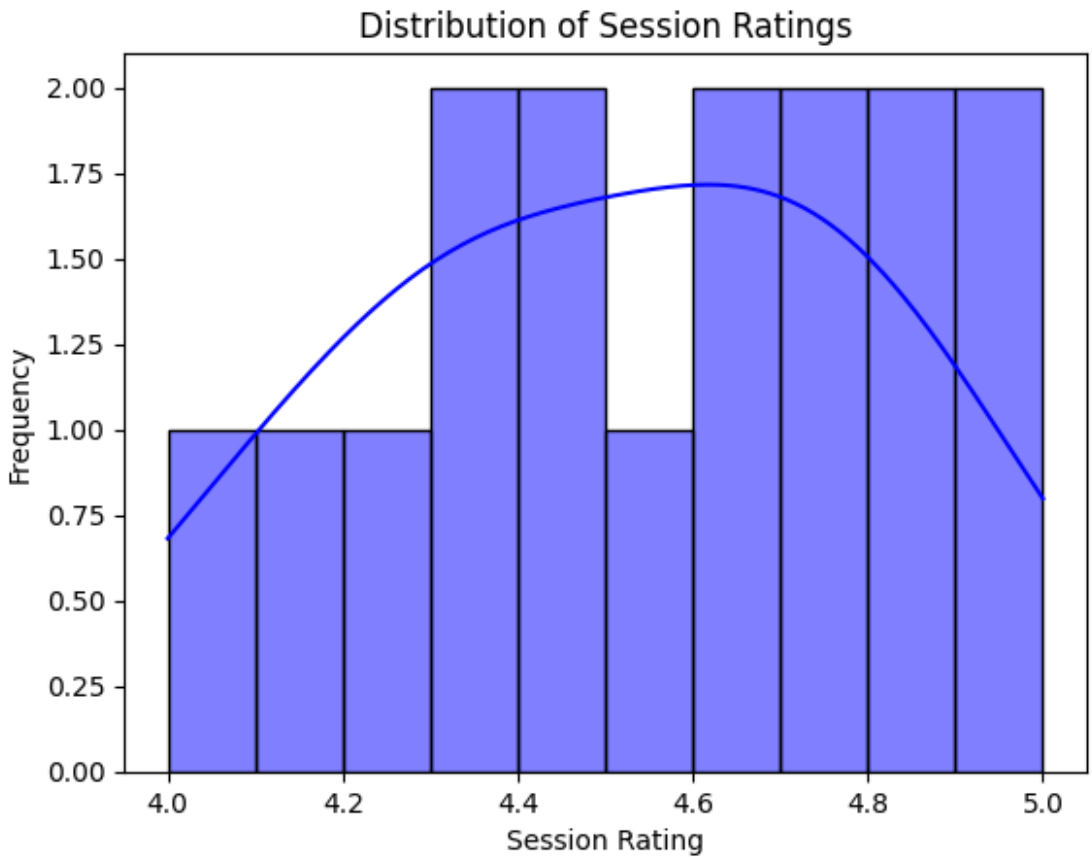


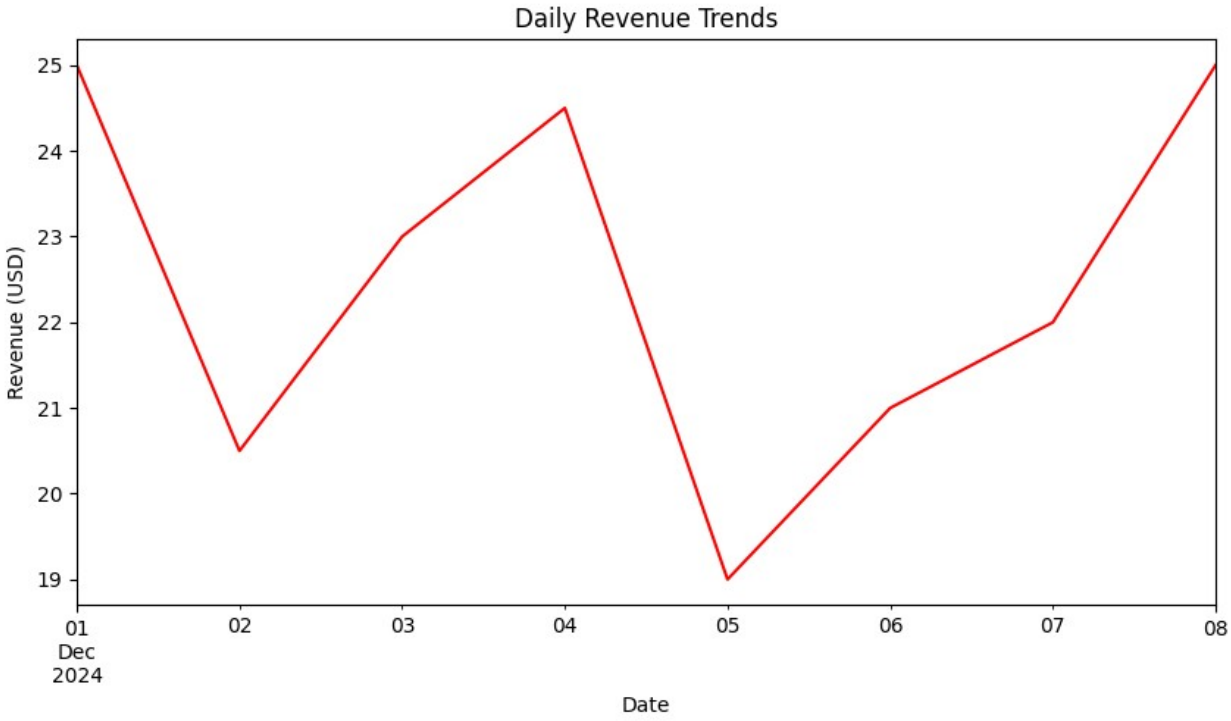

Orders per User
Number of Orders
0.0
0.5
1.0
1.5
2.0
2.5
3.0
U001
U002
U003
U004
U005
U006
U007
U008
User ID


Top 10 Popular Dishes in Cooking Sessions
Number of Sessions
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Spaghetti
Grilled Chicken
Caesar Salad
Pancakes
Veggie Burger
Oatmeal
Dish Name


Top 10 Popular Dishes in Orders
Number of Orders
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Spaghetti
Grilled Chicken
Caesar Salad
Pancakes
Veggie Burger
Oatmeal
Dish Name


Favorite Meal by Age Group
42.5
40.0
37.5
35.0
32.5
30.0
27.5
25.0
Age
Dinner
Lunch
Breakfast
Favorite Meal


Distribution of Session Ratings
Frequency
Session Rating
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
4.0
4.2
4.4
4.6
4.8
5.0


Daily Revenue Trends
Revenue (USD)
Date
19
20
21
22
23
24
25
01
Dec
2024
02
03
04
05
06
07
08

```
Analysis and visualizations completed. Summary report saved as
'Summary_Report.txt'.
```