Upliance.ai. Data analytics intern report

Task 1 - cleaning and merging the data

```
#Removing null values data from three dataset
user_details.dropna(inplace=True)
cooking_sessions.dropna(inplace=True)
order_details.dropna(inplace=True)
```

```
# Merge UserDetails and CookingSessions on 'User ID'
merged_user_cooking = pd.merge(user_details, cooking_sessions, on='User ID', how='inner', suffixes=('', '_cooking'))

# Merge the result with OrderDetails on 'Session ID'
df = pd.merge(merged_user_cooking, order_details, on='Session ID', how='inner', suffixes=('', '_order'))
```

Task 2 Analysing the relationship between cooking sessions and user orders

```
# Filter completed orders
completed_orders = df[df['Order Status'] == 'Completed']

# Calculating average session rating and order rating for completed orders
average_session_rating_completed = completed_orders['Session Rating'].mean()
average_order_rating_completed = completed_orders['Rating'].mean()

# Calculating average session duration for completed orders
average_duration_completed = completed_orders['Duration (mins)'].mean()

# Group by Order Status to compare session ratings, order ratings, and durations
session_rating_by_order_status = df.groupby('Order Status')['Session Rating'].mean()
order_rating_by_order_status = df.groupby('Order Status')['Rating'].mean()
duration_by_order_status = df.groupby('Order Status')['Duration (mins)'].mean()

# Printing results
print("Average Session Rating for Completed Orders:", average_session_rating_completed)
print("Average Order Rating for Completed Orders:", average_order_rating_completed)
print("Average Session Duration for Completed Orders (mins):", average_duration_completed)
print("\nAverage Session Rating by Order Status:\n", session_rating_by_order_status)
print("\nAverage Order Rating by Order Status:\n", order_rating_by_order_status)
print("\nAverage Session Duration by Order Status (mins):\n", duration_by_order_status)
```

Result:

Average Session Rating for Completed Orders: 4.507142857142857

Average Order Rating for Completed Orders: 4.285714285714286

Average Session Duration for Completed Orders (mins): 30.357142857142858

Average Session Rating by Order Status:

Order Status

Completed 4.507143

Name: Session Rating, dtype: float64

Average Order Rating by Order Status:

Order Status

Completed 4.285714

Average Session Duration by Order Status (mins):

Order Status

Completed 30.357143

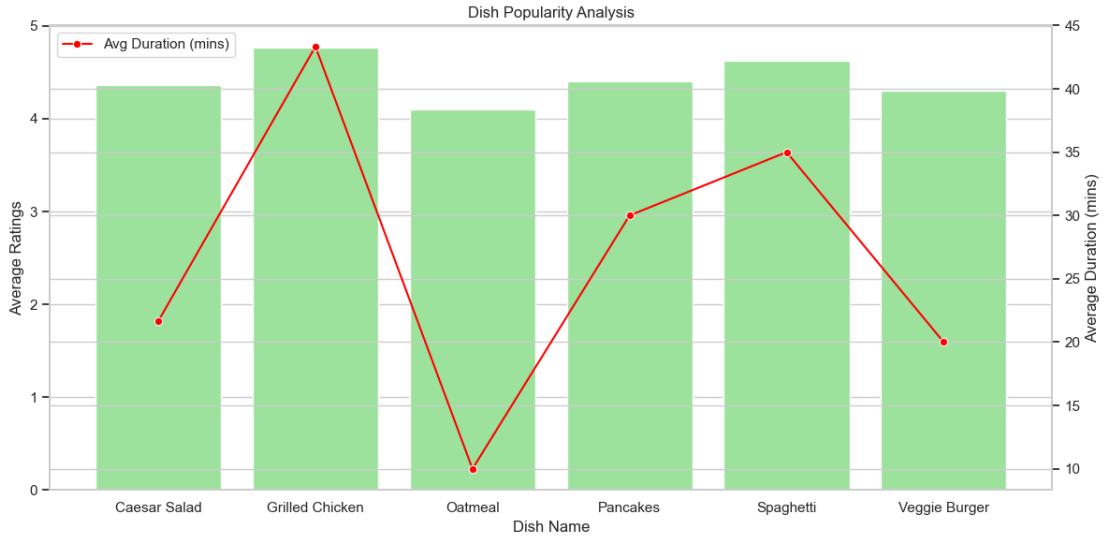
→Based on these findings I can say that the average rating 4 and average session duration has more successful orders, so to increase the sales we need to focus on the food products which has more completed orders

Task 3 - Identifying popular dishes

```
# Filtering the DataFrame for relevant columns
dish_data = df[['Dish Name', 'Rating', 'Session Rating', 'Duration (mins)']]

# Calculating average ratings and session duration for each dish
dish_analysis = dish_data.groupby('Dish Name').agg({
    'Rating': 'mean',
    'Session Rating': 'mean',
    'Duration (mins)': 'mean',
    }).reset_index()

dish_analysis.columns = ['Dish Name', 'Avg Order Rating', 'Avg Session Rating', 'Avg Duration (mins)']

# Sorting dishes by average order rating, session rating, and completion rate
popular_dishes_df = dish_analysis.sort_values(by=['Avg Order Rating', 'Avg Session Rating'], ascending=False)

# Printing the resulting DataFrame
print("Popular Dishes DataFrame:\n", popular_dishes_df)
```

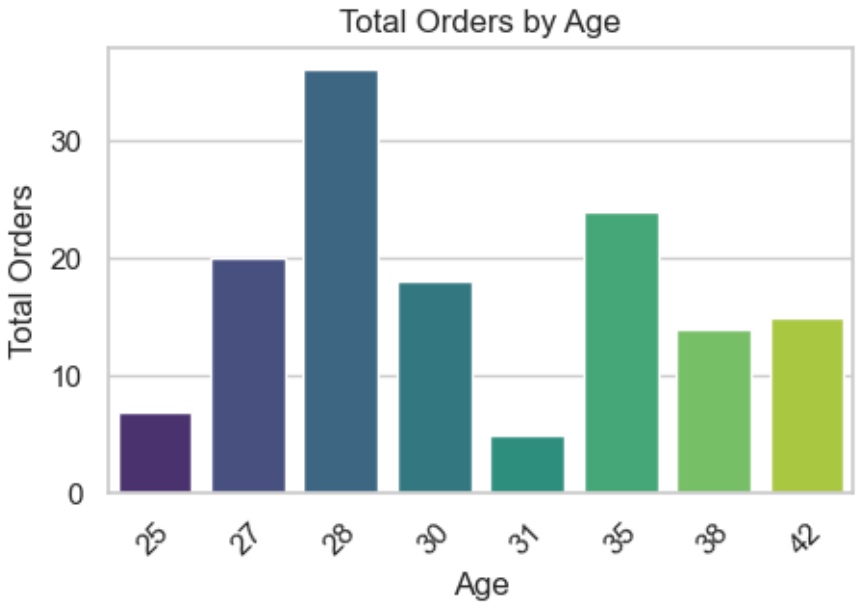
Result:

Dish Name	Avg Order Rating	Avg Session Rating	Avg Duration (mins)
Grilled Chicken	4.666666667	4.766666667	43.33333333
Spaghetti	4.5	4.625	35
Pancakes	4	4.4	30
Caesar Salad	4	4.366666667	21.66666667
Veggie Burger	4	4.3	20
Oatmeal	4	4.1	10

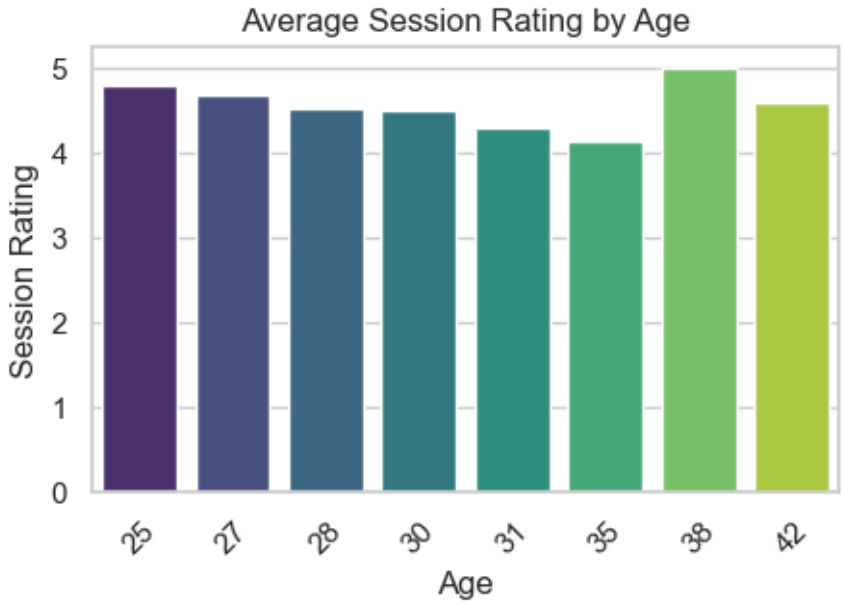


→Based on the result the most famous Dish is Grilled Chicken with average order rating and average session rating more than 4.6 and the average duration of session 43 minutes, and the second most popular is spaghetti which is almost equal to the grilled chicken and stands second

Task 4 - exploring demographic factors that influence user behaviour.



➔ *The highest orders made by people with the age 28 so focusing on the age between 25 to 35 is necessary to maintain the order level*



➔ *The session ratings seems to be good*

➔ *Based on the image we can clearly see that Miami and Austin locations are very less orders which is less then 10 so need to focus on that locations*

	Location	Total Orders	Session Rating	Amount (USD)
0	Austin	7	4.8	13
1	Boston	14	5	14
2	Chicago	15	4.6	8.5
3	Los Angeles	24	4.133333333	10.33333333
4	Miami	5	4.3	11
5	New York	36	4.533333333	11.66666667
6	San Francisco	20	4.7	10.75
7	Seattle	18	4.5	11.25

	Age	Total Orders	Session Rating	Amount (USD)
0	25	7	4.8	13
1	27	20	4.7	10.75
2	28	36	4.533333333	11.66666667
3	30	18	4.5	11.25
4	31	5	4.3	11
5	35	24	4.133333333	10.33333333
6	38	14	5	14
7	42	15	4.6	8.5