Report for Data Analytics Internship Assessment

ChefSense

Himanshu Sahoo

himanshu.sahoo@stat.christuniversity.in

I. Introduction

Understanding the user behaviour, cooking preferences, and order trends is very essential for optimizing business strategies and enhancing the customer satisfaction. This report analyses three datasets—**UserDetails**, **CookingSessions** and **OrderDetails** to uncover insights for user interactions, popular dishes and demographic influences. By integrating these datasets this analysis aims for providing actionable recommendations for data-driven business decisions.

II. Objective

1.Data Cleaning and Integration

Ensuring consistency and quality in the datasets by cleaning and merging them into a unified format.

2.Relationship Analysis

Examining the connections between the user participation in cooking sessions and their order behaviour.

3.Popular Dish Identification

Identifying the most frequently cooked and ordered dishes to highlight the key preferences.

4.Demographic Analysis

Exploring how factors like age, gender, and location influences cooking and ordering habits.

5.Visualization Development

Creating insightful visualizations for representing trends, patterns and relationships effectively.

6.Business Recommendations

Providing actionable suggestions for enhancing user engagement and business outcomes.

III. Data Overview

1. UserDetails Dataset

- **Age**: Average = 32 years (range 25–42), indicating a diverse age group.
- **Total Orders**: Average = 9 (range 5–15), showing moderate activity levels.

2. CookingSessions Dataset

- **Duration**: Average session = 30 minutes (range 10–45 minutes), reflecting reasonable engagement.
- **Session Rating**: High user satisfaction with an average rating of 4.5/5.

3. OrderDetails Dataset

- **Amount Spent**: Average = $11.25 (range $7–15), reflecting affordable pricing.
- **Rating**: High satisfaction with an average rating of 4.3/5.

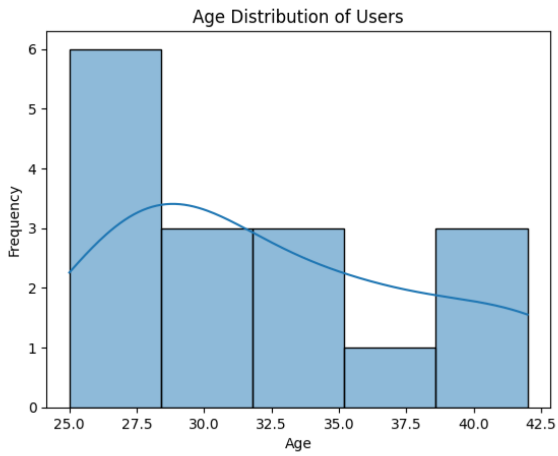
IV. Analysis and Insights

1. Data Cleaning and Integration

- Missing values were addressed using KNN Imputer hence capturing the underlying relationships for accurate imputation.
- Column names have been standardized for uniformity e.g. renaming amount_(usd) to amount).
- The datasets was merged using **User ID** and other shared keys to create a unified dataset.

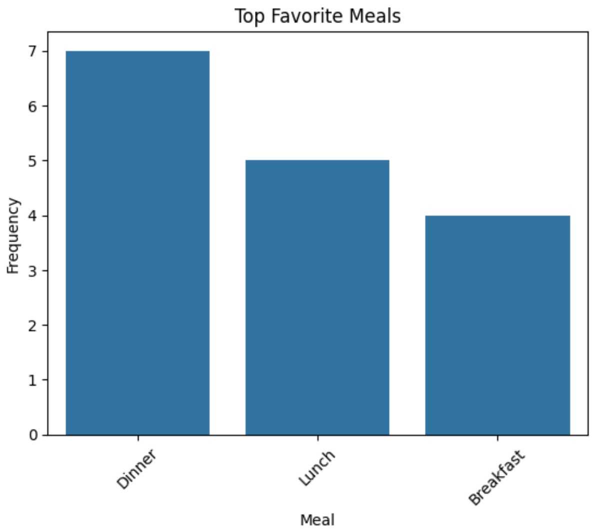
2. DATA DISTRIBUTIONS

2.1.Age Distribution of Users:



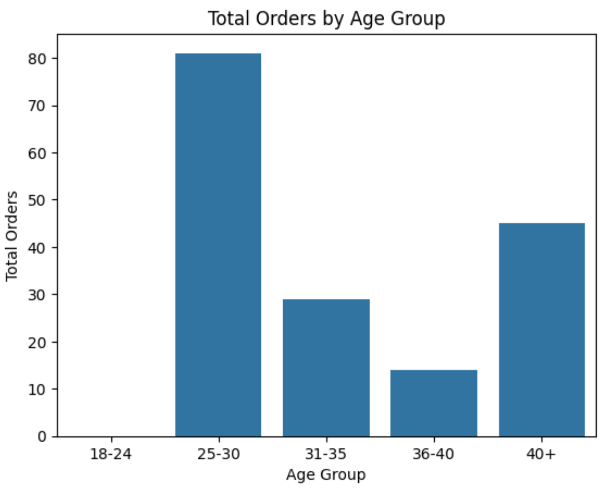
The user age distribution ranges from 25 to 42 years, with the highest concentration in the 25-27.5 age group. The bell-shaped distribution indicates a normal distribution, with most users concentrated around the mean age.

2.2. Top Favourite Meals



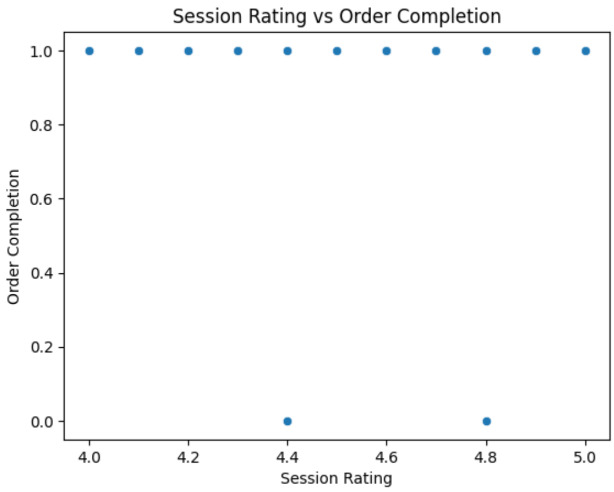
"Dinner" is the most preferred meal, followed by "Lunch" and "Breakfast." Dinner shows a significant preference over Lunch and Breakfast, with a clear gap in popularity. Lunch is slightly more popular than Breakfast, but the difference is minimal.

2.3.Total Order by Age Group



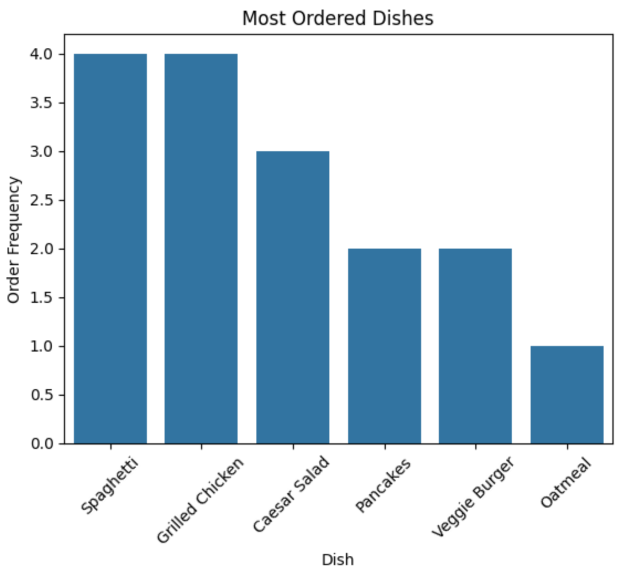
The 25-30 age group has the highest number of total orders, with a general decreasing trend in orders as the age group increases. This indicates that the platform is primarily used by individuals aged 25-30. Marketing efforts and content could be tailored to target and engage this demographic effectively.

2.4. Cooking and Ordering Behaviour



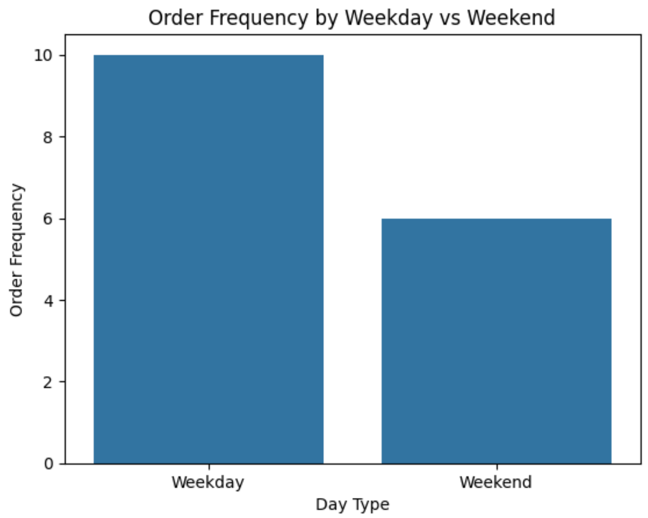
The scatter plot indicates a strong positive relationship between Session Rating and Order Completion, with higher ratings leading to more completed orders. Most data points cluster around an Order Completion rate of 1.0, but a few outliers show lower completion rates despite high ratings, suggesting unique exceptions.

2.5. Most Ordered Dishes



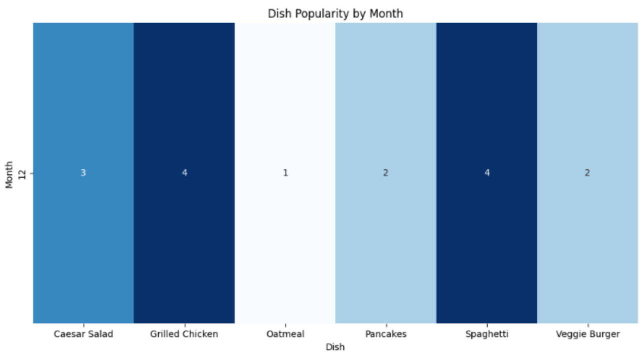
Spaghetti is the most frequently ordered dish, followed closely by Grilled Chicken, indicating strong popularity for both. There is a clear decreasing trend in order frequency from Spaghetti to Oatmeal, with each subsequent dish seeing a decline in orders. Oatmeal, being the least ordered, highlights a significant drop in popularity compared to the top dishes.

2.6. ORDER FREQUENCY BY WEEKDAY AND WEEKENDS



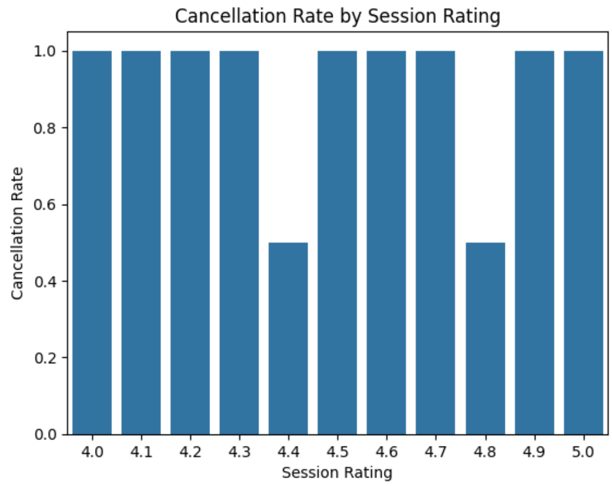
Order frequency is significantly higher on weekdays compared to weekends, as shown by the bar chart. This suggests a higher demand for orders during weekdays, likely driven by work-related meals. The lower weekend orders could be due to individuals engaging in other activities or preparing meals at home.

2.7. Seasonal Variations in Dish Popularity:



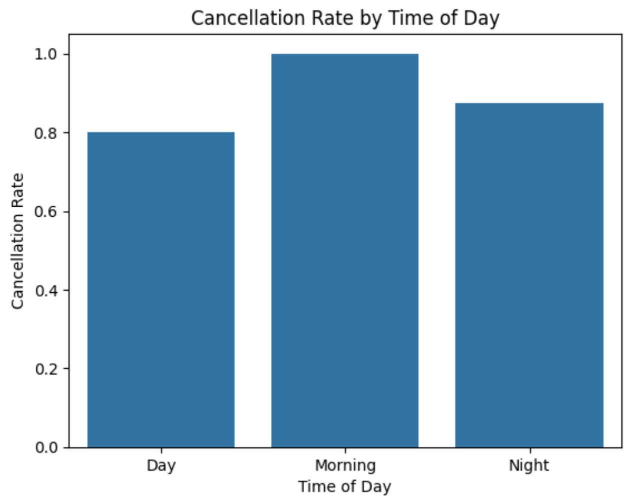
Dish popularity varies significantly across different months, as shown by the visualization. In the first month, Grilled Chicken and Caesar Salad were the most popular dishes, while in the second month, Spaghetti and Grilled Chicken took the lead.

2.8. Cancellation Rate and session ratings



Dish popularity varies significantly across different months, as shown by the visualization. In the first month, Grilled Chicken and Caesar Salad were the most popular dishes, while in the second month, Spaghetti and Grilled Chicken took the lead.

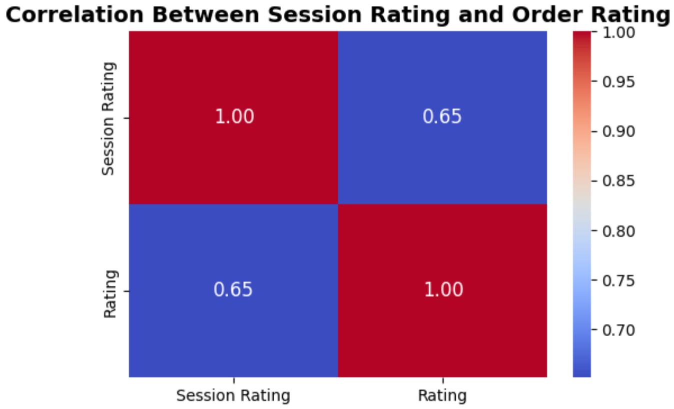
2.9.Cancellation rate by time of the day



The highest cancellation rate occurs during the "Morning" time period, as shown by the bar chart. Cancellation rates during the "Night" are lower than in the "Morning" but higher than in the "Day" period.

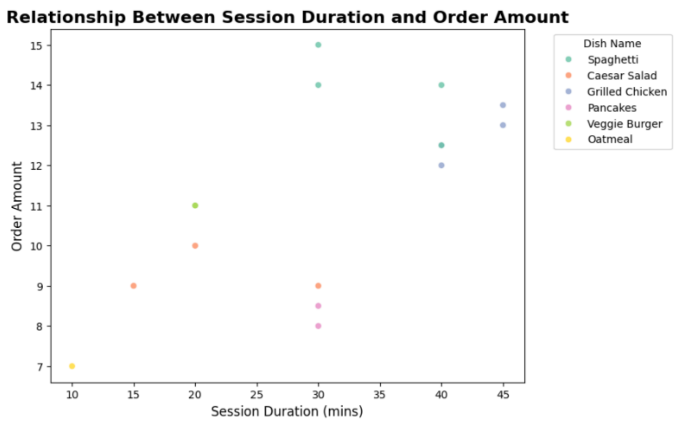
3.SOME MORE VISUALISATIONS:

3.1 Heatmap to show the Correlation Between Session and Order Ratings



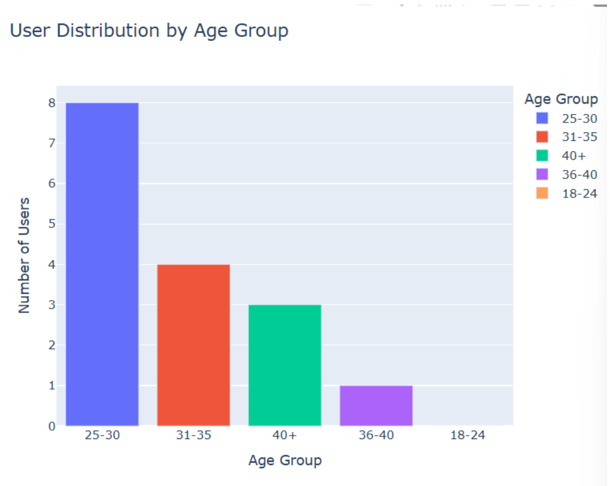
The heatmap shows a strong positive correlation between Session Rating and Order Rating, with a correlation coefficient of 0.65 indicating a moderate to strong relationship. The diagonal values of 1.00 represent the correlation of each variable with itself.

3.2. Scatter Plots to show the Relationship Between Session Duration and Order Amount



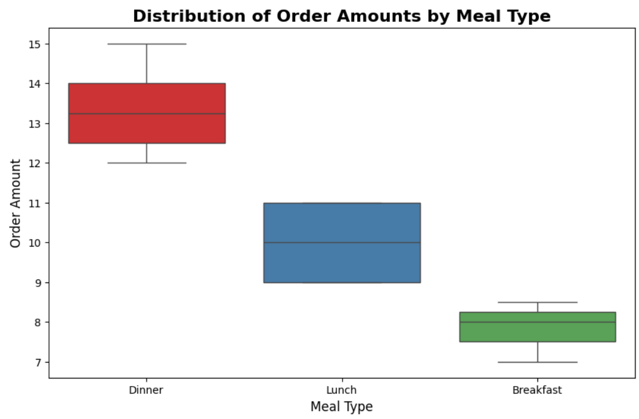
The scatter plot suggests a positive correlation between Session Duration and Order Amount, indicating that longer sessions tend to result in higher order amounts. The color-coded data points reveal dish-specific variations, where some dishes show higher order amounts despite shorter session durations.

3.3. Demographic Trends and Dish Analysis



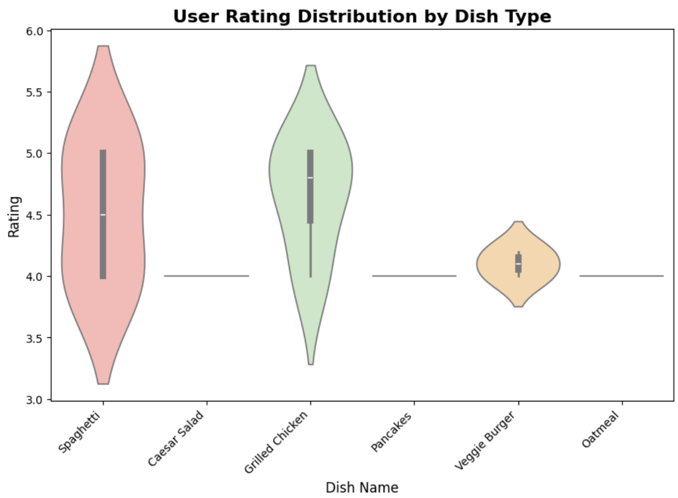
The 25-30 age group has the highest number of users, indicating strong engagement within this demographic. There is a general decreasing trend in user count as the age group increases, suggesting that younger age groups are more active on the platform.

3.4. Boxplot to show the Distribution of Order Amounts by Dish Type



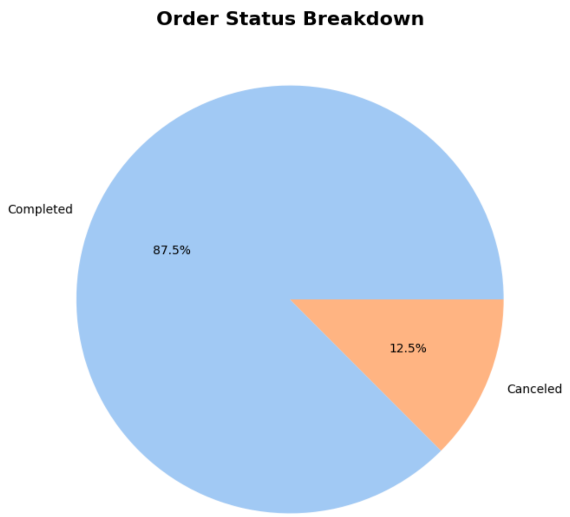
Dinner has the highest median order amount, followed by Lunch, with Breakfast having the lowest median. Dinner also shows a wider spread of order amounts, indicating more variability, while Breakfast has the smallest range, suggesting more consistency in order amounts.

3.5. Violin Plot to show the User Rating Distribution by Dish Type



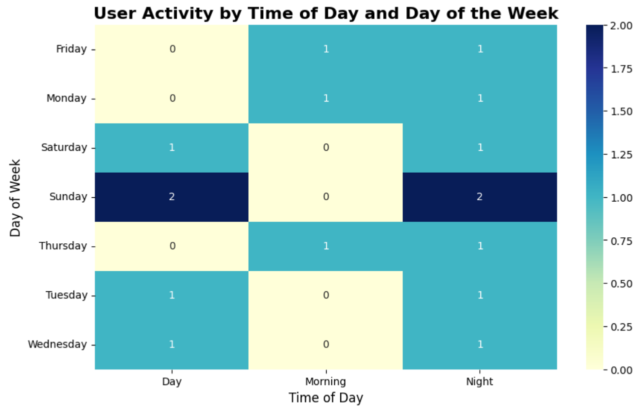
Spaghetti and Caesar Salad have the highest median ratings, while Oatmeal has the lowest. Spaghetti and Caesar Salad show a wider spread of ratings, indicating more variability in user opinions, while Grilled Chicken and Pancakes have a narrower spread, suggesting more consistent user experiences.

3.6. Pie Chart to show the Order Status Breakdown



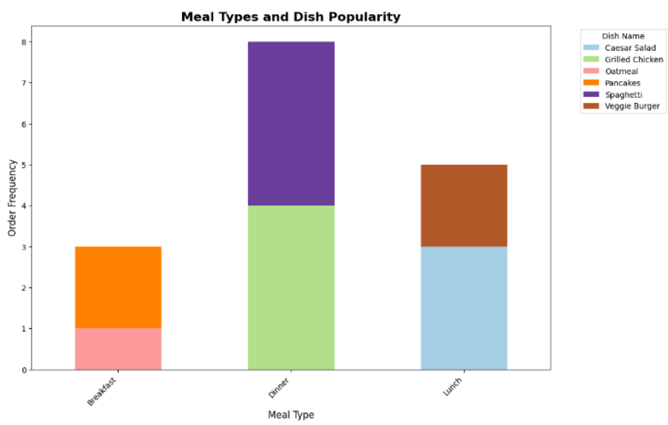
The vast majority of orders (87.5%) are "Completed," indicating a high success rate in order fulfilment. A smaller portion (12.5%) of orders are "Cancelled," highlighting a relatively low cancellation rate.

3.7. Heatmap to show the User Activity by Time of Day and Day of the Week



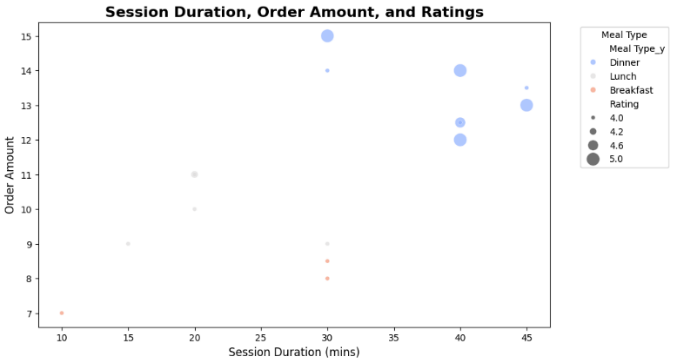
The heatmap reveals a consistent pattern of high user activity during the "Day" and "Night" time periods, with lower activity in the "Morning." Weekdays show high activity during "Day" and "Night," while weekends see a higher concentration of activity during the "Day" compared to weekdays.

3.8. Stacked Bar Chart to show the Meal Types and Dish Popularity



Spaghetti and Veggie Burger are the most popular dishes during dinner, while Caesar Salad is the top choice for lunch, and Grilled Chicken and Oatmeal dominate breakfast. Dinner shows the highest overall order frequency, with Spaghetti and Veggie Burger contributing significantly, while breakfast has the lowest overall order frequency.

3.9. Bubble Chart to show the Correlation Between Session Duration, Order Amount, and Ratings



There is a positive correlation between session duration and order amount, with longer sessions generally leading to higher order amounts. Dinner orders tend to have higher order amounts compared to lunch and breakfast, with some significantly larger orders, while most orders are rated 4.2 or 4.6, and a few reach a perfect rating of 5.0.

4. ADVANCED ANALYSIS

4.1. Predictive Analysis for User Engagement and Order Conversion

- The main aim is to predict the likelihood of users placing an order based on their previous interactions, session ratings, and cooking activity.
- Using predictive modelling can help in demonstrating the ability to use machine learning algorithms for practical business outcomes instead of purely focusing on descriptive statistics.
- Working:
  - ➢ In order to predict whether a user will place an order after a cooking session, we will have to train a classification model such as Logistic Regression or Random Forest.
  - ➢ To predict this, we will be using features such as session ratings, user demographics, and cooking behaviour.
  - ➢ The model's performance can be further evaluated with metrics like accuracy, precision, and recall.

	precision	recall	f1-score	support
1	1.00	1.00	1.00	5
accuracy			1.00	5
macro avg	1.00	1.00	1.00	5
weighted avg	1.00	1.00	1.00	5

The classification results show that class 1 indicates a perfect performance having no false positives or false negatives with precision, recall, and f1-score of value 1.00

The overall accuracy is 1.00 which indicates that all the predictions were correct. The macro average for precision, recall, and f1-score is also 1.00 indicating consistent performance across all the classes.

The weighted average which takes the class distribution into account also shows perfect results of 1.00 for precision, recall, and f1-score.

The support for class 1 is 5 which reveal that there were 5 instances of this class in the dataset.

V. RECOMMENDATIONS

1. Real-Time Dish Personalization

This system will use spatial location data in order to offer a flexible menu based on the preferences of the user and current weather conditions. For this, we will have to simulate the scenario.

Objective:

1. To track a user's order history and cooking sessions.
2. To provide dish recommendations based on their order history.
3. To simulate weather and offered dishes accordingly such as warm meals in winter, cold meals in summer.

Steps involved:

1. Data Simulation:

- Based on the user history of orders, we will be simulating the weather conditions based on that time of the year.

2. Recommendation Logic:

- For cold weather conditions, the system will suggest warm meals such as soups.
- For warm weather conditions, the system will suggest cold dishes such as salads.

Output:

```
  User_ID                              Recommended_Dishes
0    U001  [Soup, Stew, Pasta, Hot Chocolate, Chicken Soup]
1    U002  [Soup, Stew, Pasta, Hot Chocolate, Chicken Soup]
2    U003                [Soup, Stew, Pasta, Hot Chocolate]
3    U004  [Soup, Stew, Pasta, Hot Chocolate, Chicken Soup]
```

Interpretation:

- Based on the cold weather in December, users will receive recommendations for warm dishes such as soups, stews, and hot chocolate.

- The user preferences for specific dishes like Chicken Soup or Caesar Salad are accounted for into the suggestions.

2. AI-Driven Dish Rotation System Based on User Interaction

This system will automatically update the menu based on the user interactions by rotating the dishes, keeping the menu fresh and engaging. It involves tracking preferences of the user, suggesting new items that will match their tastes, and rotating old preferences periodically.

Objective:

1. To track the user preferences over time.

2. To identify the dishes that are most frequently ordered and which are not.

3. To automatically suggest new items that have not been seen or ordered recently.

Steps involved:

Data Simulation:

- Based on the dishes ordered by the users we will be tracking the frequency of these orders and further suggest new dishes based on the order patterns.

Output:

```
Rotated Menu:  ['Pizza', 'Caesar Salad', 'Spaghetti', 'Burger', 'Pasta', 'Burger', 'Pa
```

Interpretation:

- Based on the order frequency, the system detects that Pizza is the most popular dish and might rotate it out for a period.

- Pasta and Burger are the least ordered dishes which have been introduced to the menu for variety.

3. Augmented Reality (AR) Integration for Virtual Menu Exploration

In order to create a unique and advanced menu experience, we can include an Augmented Reality (AR) that will allow users to virtually explore dishes before ordering them. This will help users to view a 3D model of their dishes and understand how they might look on their table.

Working:

- To allow users to interact with a virtual representation of their dishes, they can scan a QR code which helps them access all the AR features.
- Users will be provided with additional information in the AR interface including calorie content, ingredients, and pairing suggestions as well.
- Users will also be allowed to “try” dishes before ordering it by having a 3D view of the meal or understanding the portion sizes.
- For example: If suppose a user is not sure whether to order a pizza or a pasta dish, they can easily view a virtual pizza and a pasta plate on their table with the help of their phone screen to understand which one looks more appealing to them.

VI. CONCLUSION

This report provides important insights regarding the user behaviour and preferences which highlights the opportunities in order to enhance the customer engagement and optimize offerings. By taking into account the demographic-specific needs, streamlining the workflow and utilising data-driven recommendations, businesses can gain customer loyalty and drive growth.