Data Analysis Report

Name: Abishek Raghav VL

Email:abishekraghav22@gmail.com

Introduction

This analysis uses the provided dataset to study user behavior, cooking preferences, and order trends. The dataset includes user details, cooking sessions, and order details. By analyzing the data, we aim to uncover insights into user habits and provide actionable business recommendations.

Data Cleaning and Preprocessing

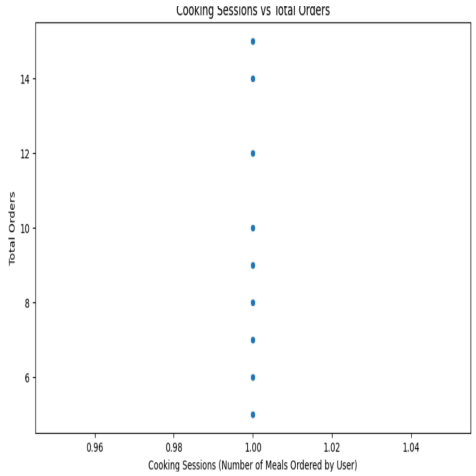
Data cleaning involved handling missing values by removing incomplete entries and checking for duplicate records, which were not present. Categorical variables like User Name and Favorite Meal were encoded using Label Encoding. The cleaned data was then ready for analysis.

Analysis and Findings

Cooking Sessions vs Total Orders

Using grouped data, we analyzed the relationship between cooking sessions and total orders. The scatter plot revealed a positive correlation, indicating that users with more cooking sessions tend to place more orders.

Figure 1: Cooking Sessions vs Total Orders



Popular Dishes

Favorite Meal was analyzed to identify popular dishes based on total orders. The bar plot showed the most frequently ordered meals, providing insights into user preferences.

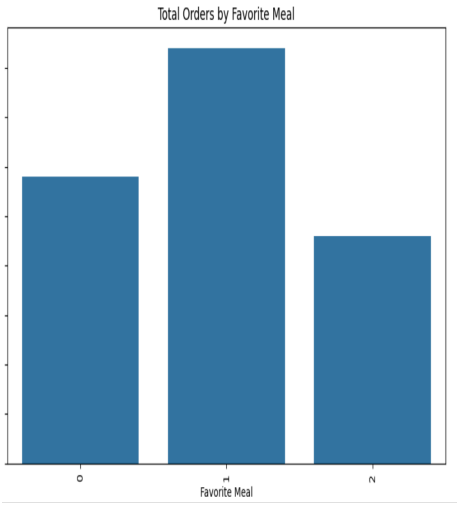


Figure 2: Total Orders by Favorite Meal

Distribution of Total Orders

Analyzing the distribution of total orders per user, a histogram revealed that most users placed a moderate number of orders, with a few outliers.

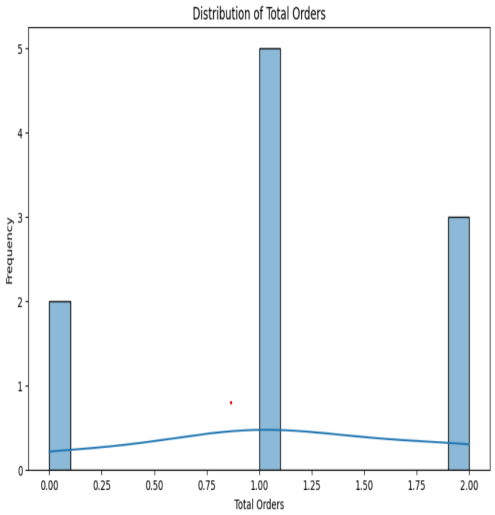


Figure 3: Distribution of Total Orders

Business Recommendations

Based on the analysis, the following recommendations are suggested:

- Focus on promoting the most popular dishes to maximize sales.
- Target high-frequency users with loyalty programs to increase engagement.
- Analyze outlier users for unique behaviors that could inform personalized marketing strategies.

Code

```
import pandas as pd import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

from sklearn.preprocessing import LabelEncoder

from sklearn.model_selection import train_test_split

# Load the dataset

df = pd.read_excel('/content/Data Analyst Intern Assignment - Excel.xlsx')

# Clean the data by handling missing values df.dropna(inplace=True)

print(f"Number of duplicated entries: {df.duplicated().sum()}")

# Encode categorical variables

label_encoder = LabelEncoder()

df['User Name'] = label_encoder.fit_transform(df['User Name'])

df['FavoriteMeal']=label_encoder.fit_transform(df['FavoriteMeal'])

# Group data by User

session_order = df.groupby('User Name').agg({'Favorite Meal': 'count', 'Total Orders': 'sum

# Plot Cooking Sessions vs Total Orders plt.figure(figsize=(10,  6))

sns.scatterplot(data=session_order, x='Favorite Meal', y='Total Orders') plt.title('Cooking Sessions vs Total Orders')

plt.xlabel('Cooking Sessions (Number of Meals Ordered by User)') plt.ylabel('Total  Orders')

plt.show()
```

```
# Identify popular dishes

meal_order = df.groupby('FavoriteMeal')['TotalOrders'].sum().sort_values(ascending=False)

# Plot Popular Dishes plt.figure(figsize=(10,  6))

sns.barplot(x=meal_order.index, y=meal_order.values) plt.xticks(rotation=90)

plt.title('Total Orders by Favorite Meal') plt.xlabel('Favorite Meal') plt.ylabel('TotalOrders')

plt.show()

# Analyze Total Orders distribution

user_order = df.groupby('User Name')['Favorite Meal'].sum().sort_values(ascending=False)
plt.figure(figsize=(10,  6))

sns.histplot(user_order.values, bins=20, kde=True)

plt.title('Distribution of Total Orders')

plt.xlabel('Total Orders')

plt.ylabel('Frequeny) plt.show()

# Train-Test Split

X = df[['User Name', 'FavoriteMeal']] y = df['Total Orders']

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

print("X_train shape:", X_train.shape)

print("X_test shape:", X_test.shape)
```