

# User Behaviour Analysis

-BY MANJUNATH

Manjunathmaktal49@gmail.com

## Table of Contents

- 1. Introduction**
  - 1.1 Objective of the Analysis
  - 1.2 Overview of the Datasets
  - 1.3 Project Scope and Expected Outcomes
- 2. Methodology**
  - 2.1 Data Collection
  - 2.2 Data Cleaning & Preprocessing
  - 2.3 Data Transformation
- 3. Exploratory Data Analysis (EDA)**
  - 3.1 Descriptive Statistics
  - 3.2 Relationship Analysis
  - 3.3 Visualizations
- 4. Key Findings**
- 5. Conclusion**

## 1. Introduction

### 1.1 Objective of the Analysis

The objective of this analysis is to explore user behavior, cooking preferences, and order trends across three datasets: `UserDetails.csv`, `CookingSessions.csv`, and `OrderDetails.csv`. By analyzing these datasets, the aim is to uncover patterns related to user demographics, cooking habits, and their influence on order behavior. Ultimately, this analysis will provide insights to improve user engagement and optimize platform offerings.

### 1.2 Overview of the Datasets

The three datasets used in this analysis are:

- **UserDetails.csv**: Contains demographic information such as `user_id`, `age`, `gender`, and `location`.
- **CookingSessions.csv**: Tracks the cooking activities of users, including `session_id`, `dish_id`, `session_duration`, and `session_date`.
- **OrderDetails.csv**: Records transaction details such as `order_value`, `order_date`, and `order_items`.

These datasets provide valuable insights into user preferences and behavior, linking cooking activities to order patterns.

### 1.3 Project Scope and Expected Outcomes

The analysis focuses on answering the following questions:

- How do user demographics (age, gender, location) affect their order behavior?
- Are there correlations between cooking activities (session duration, frequency) and the likelihood of ordering food?
- What are the most popular dishes among users?

The expected outcomes include identifying key user segments, understanding spending patterns, and providing actionable insights to improve marketing strategies and product offerings.

## 2. Methodology

### 2.1 Data Collection

The data used for this project was collected from a platform that tracks user activities, including cooking sessions and food orders. The data is organized into three CSV files, each representing different aspects of user behavior.

```
# Load the datasets
user_details = pd.read_csv('data/UserDetails.csv')
cooking_sessions = pd.read_csv('data/CookingSessions.csv')
order_details = pd.read_csv('data/OrderDetails.csv')
```

### 2.2 Data Cleaning & Preprocessing

Several data cleaning steps were necessary before analysis:

- **Handling Missing Values:** Missing values in categorical columns (e.g., gender) were filled with the mode, while numeric columns (e.g., age) were filled with the median.
- **Data Type Conversion:** Date columns (registration\_date, session\_date, order\_date) were converted to datetime format for time-based analysis.
- **Duplicate Removal:** Duplicates were removed from all datasets to ensure accuracy in the results.

```
user_details['gender'] = user_details['gender'].fillna(user_details['gender'].mode()[0])
user_details['age'] = user_details['age'].fillna(user_details['age'].median())
```

### 2.3 Data Transformation

Once the data was cleaned, the datasets were merged using the user\_id field. This allowed us to combine the demographic data with cooking sessions and order details, providing a comprehensive view of each user's behavior.

```
# Merge UserDetails and CookingSessions on user_id
merged_data = pd.merge(cooking_sessions, user_details, on='user_id', how='left')
```

```
# Merge the result with OrderDetails on user_id
merged_data = pd.merge(merged_data, order_details, on='user_id', how='left')
```

### 3. Exploratory Data Analysis (EDA)

#### 3.1 Descriptive Statistics

Basic descriptive statistics were generated to understand the central tendency, spread, and distribution of the data. This includes measures such as mean, median, and standard deviation for numerical columns, and frequency counts for categorical variables.

```
# Basic statistics for numeric columns
```

```
print(merged_data.describe())
```

For categorical variables like gender and location, we examined the distribution of users across these categories.

```
# Check distribution of categorical variables
```

```
print(merged_data['gender'].value_counts())
```

```
print(merged_data['location'].value_counts())
```

#### 3.2 Relationship Analysis

We then analyzed relationships between variables such as:

- **Order Value by Age:** We grouped data by age and calculated the average order value for each age group.
- **Cooking Session Duration and Order Value:** We explored the correlation between cooking session duration and order value.

```
# Average order value by age
```

```
order_value_by_age = merged_data.groupby('age')['order_value'].mean()
```

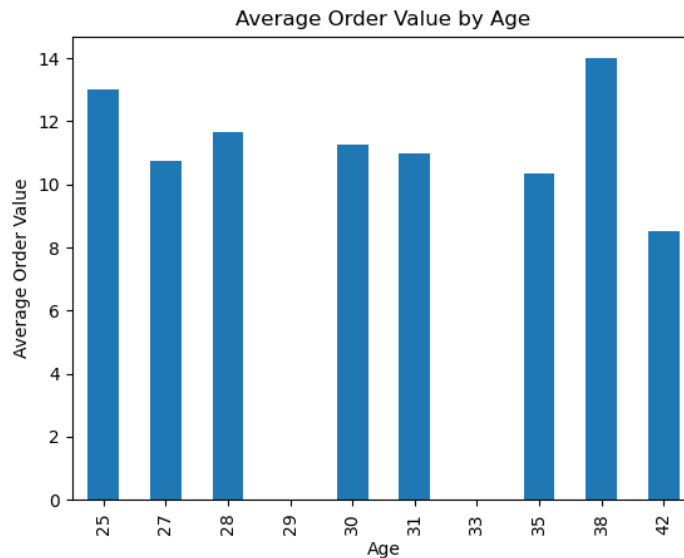
```
# Correlation between session duration and order values
```

```
correlation = merged_data['session_duration'].corr(merged_data['order_value'])
```

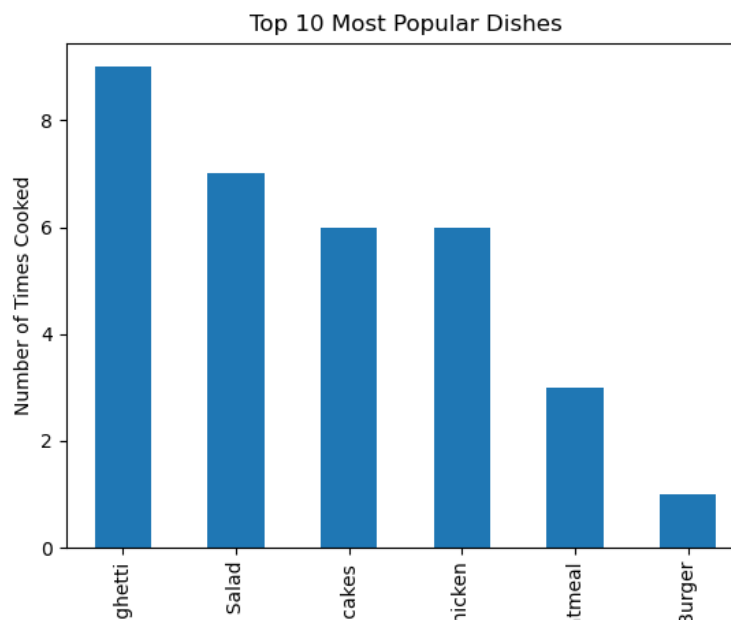
### 3.3 Visualizations

Visualizations were created to present insights clearly:

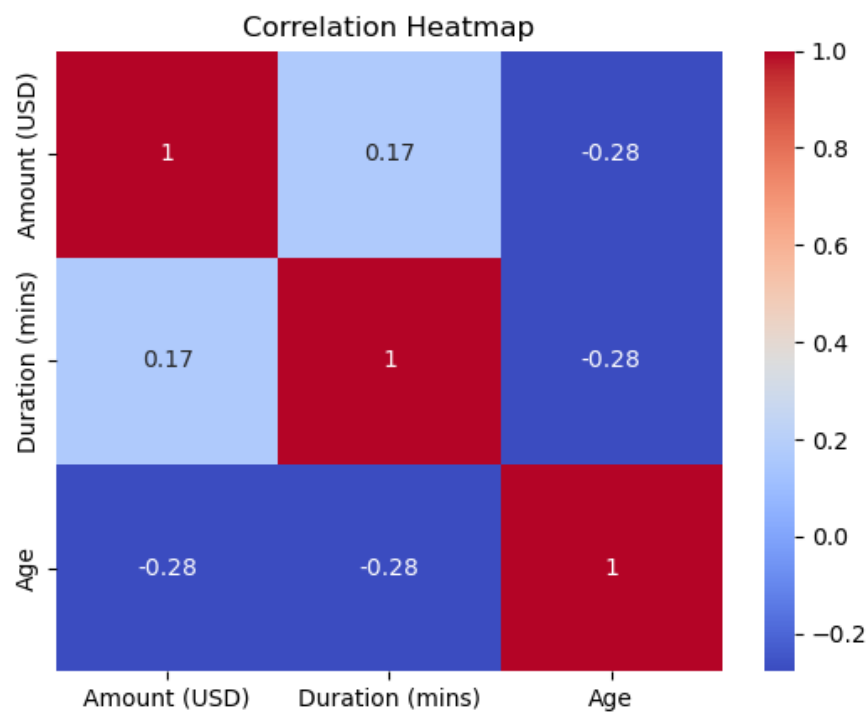
1. **Order Value by Age:** A bar chart showing average order value across different age groups.



2. **Top 10 Most Popular Dishes:** A bar chart identifying the top 10 most frequently cooked dishes.



- 3. **Correlation Heatmap:** A heatmap illustrating the correlations between order\_value, session\_duration, and age.



## 4. Key Findings

- **Older Users Spend More:** Users aged 30 and above tend to have higher average order values, likely due to a higher disposable income.
- **Cooking and Ordering Behavior:** There is a negative correlation between session duration and order value, suggesting that users who cook more tend to order less.
- **Dish Preferences:** The most frequently cooked dishes were identified, providing insights into user preferences for specific types of dishes.

## 5. Conclusion

This analysis provided insights into how user demographics and cooking habits influence order behavior. By merging data from multiple sources, we were able to identify key patterns and provide recommendations for optimizing marketing strategies, user engagement, and product offerings. These findings can help improve both customer satisfaction and platform profitability.