**Data Analysis**

**and**

**Visualization**

**CA1**

**Specification Index Generation**

**and**

**Visualization**

**Documentation**

**Name:** CHIA HOU TAN

**ID:** D00223937

**Group:** SD4B

**Course:** Computing In Software Development Level 8

**GitHub:** https://github.com/ChiaHouTan/Data-Analysis-Visualisation-CA1.git

Table of Contents

[Theoretical Framework 4](#__RefHeading___Toc107_1255659565)

[Data Selection 5](#__RefHeading___Toc113_1255659565)

[Data 5](#__RefHeading___Toc122_1255659565)

[Description 5](#__RefHeading___Toc115_1255659565)

[Attributes 5](#__RefHeading___Toc117_1255659565)

[Purpose 5](#__RefHeading___Toc119_1255659565)

[Head of Dataset 6](#__RefHeading___Toc124_1255659565)

[Shape of Dataset 6](#__RefHeading___Toc126_1255659565)

[Info of Dataset 6](#__RefHeading___Toc128_1255659565)

[Percent Missing of Dataset 7](#__RefHeading___Toc130_1255659565)

[Null Values of Dataset 7](#__RefHeading___Toc132_1255659565)

[Numerical Summary of Dataset 7](#__RefHeading___Toc134_1255659565)

[Duplicate Values of Dataset 8](#__RefHeading___Toc136_1255659565)

[Duplicate Values of Dataset after DROP 8](#__RefHeading___Toc138_1255659565)

[Multivariate Analysis 9](#__RefHeading___Toc140_1255659565)

[Plot Distribution of Age 9](#__RefHeading___Toc142_1255659565)

[Plot Distribution of Gender 10](#__RefHeading___Toc142_1255659565_Copy_1)

[Plot Distribution of Marital Status 11](#__RefHeading___Toc142_1255659565_Copy_1_)

[Plot Distribution of Occupation 12](#__RefHeading___Toc142_1255659565_Copy_11)

[Plot Distribution of Monthly Income 13](#__RefHeading___Toc142_1255659565_Copy_12)

[Plot Distribution of Educational Qualifications 14](#__RefHeading___Toc142_1255659565_Copy_13)

[Plot Distribution of Family Size 15](#__RefHeading___Toc142_1255659565_Copy_14)

[Plot Distribution of Output 16](#__RefHeading___Toc142_1255659565_Copy_15)

[Plot Distribution of Feedback 17](#__RefHeading___Toc142_1255659565_Copy_16)

[PCA - Principal Component Analysis 18](#__RefHeading___Toc169_1255659565)

[Normalization 18](#__RefHeading___Toc171_1255659565)

[Explained Variance Ratio 18](#__RefHeading___Toc173_1255659565)

[Explained Variance Ratio by Principal Component 18](#__RefHeading___Toc175_1255659565)

[Component Loading 19](#__RefHeading___Toc177_1255659565)

[Plot of Online Food Ordering Dataset 19](#__RefHeading___Toc179_1255659565)

[Plot of Online Food Ordering Dataset with Clusters 20](#__RefHeading___Toc181_1255659565)

[Multiple Regression 21](#__RefHeading___Toc183_1255659565)

[Coefficient and Intercept 21](#__RefHeading___Toc185_1255659565)

[PairPlot 22](#__RefHeading___Toc187_1255659565)

[Normalization 22](#__RefHeading___Toc189_1255659565)

[Cluster Analysis 23](#__RefHeading___Toc191_1255659565)

[Normalization 23](#__RefHeading___Toc193_1255659565)

[Elbow Method for Optimal Number of Clusters 23](#__RefHeading___Toc195_1255659565)

[Clusters PairPlot 24](#__RefHeading___Toc197_1255659565)

[Outlier Detection 25](#__RefHeading___Toc199_1255659565)

[Numerical Feature Distribution by Gender 26](#__RefHeading___Toc201_1255659565)

[Weighting and Aggregation 27](#__RefHeading___Toc203_1255659565)

[Weighting and Aggregation by Gender 27](#__RefHeading___Toc208_1255659565)

[Weighting and Aggregation by Age Groups 27](#__RefHeading___Toc210_1255659565)

[Weighting and Aggregation by Geographic Location 27](#__RefHeading___Toc212_1255659565)

[Weighting and Aggregation by Educational Qualifications 28](#__RefHeading___Toc214_1255659565)

[Weighting and Aggregation by Monthly Income 28](#__RefHeading___Toc216_1255659565)

[Weighting and Aggregation by Occupation 28](#__RefHeading___Toc218_1255659565)

[Link to other indices 29](#__RefHeading___Toc220_1255659565)

[Hypothetical Composite Index: Online Food Ordering Experience Index (OFOEI) 29](#__RefHeading___Toc222_1255659565)

[Comparing OFOEI with Established Indices 29](#__RefHeading___Toc224_1255659565)

[Comparison with Net Promoter Score (NPS) for Service Industries 29](#__RefHeading___Toc226_1255659565)

[Comparison with E-commerce User Experience Indexes 29](#__RefHeading___Toc228_1255659565)

# Theoretical Framework

In the rapidly evolving landscape of online food ordering platforms, stakeholders are increasingly interested in understanding the effectiveness and impact of various platform features and services on user satisfaction and engagement. To address this need, our team aims to develop a composite index that provides a comprehensive measure of user experience and platform performance within the online food ordering sector.

**Relevance to Stakeholders:** Online food ordering platforms are facing intense competition, and stakeholders, including platform operators, policymakers, and investors, require robust metrics to assess and benchmark platform performance. A composite index that captures multiple dimensions of user experience and platform performance can serve as a valuable tool for stakeholders to make informed decisions and prioritize strategic initiatives.

**Complexity of User Behaviour:** User behaviour in the online food ordering sector is multifaceted and influenced by various factors such as demographic characteristics, socio-economic status, and geographic location. A composite index allows us to integrate these diverse dimensions into a single measure, providing a holistic view of user experience and platform performance.

**Appropriateness of Data:** The dataset contains a rich set of variables, including age, gender, marital status, occupation, monthly income, educational qualifications, family size, geographic location, and user feedback. These variables reflect key aspects of user demographics, socio-economic status, and behaviour, making them appropriate for constructing a composite index that captures the complexity of user experience and platform performance.

**Potential for Insights and Actionable Recommendations:** By analysing the relationships between different variables in the dataset and constructing composite indices, we can uncover insights into the factors that drive user satisfaction and engagement on online food ordering platforms. These insights can inform actionable recommendations for platform operators to enhance user experience, optimize service offerings, and differentiate themselves in the competitive market landscape.

Overall, developing a composite index based on the provided dataset enables us to address the evolving needs of stakeholders in the online food ordering sector by providing a comprehensive and actionable measure of user experience and platform performance.

# Data Selection

## Data

**onlinefoods.csv**

## Description

The dataset contains information collected from an online food ordering platform over a period of time. It encompasses various attributes related to Age, Gender, Marital Status, Occupation, Monthly Income, Educational Qualifications, Family Size, Latitude, Longitude, Pin Code, Output and Feedback.

## Attributes

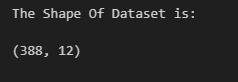
* **Age:** Age of the customer.
* **Gender:** Gender of the customer.
* **Marital Status:** Marital status of the customer.
* **Occupation:** Occupation of the customer.
* **Monthly Income:** Monthly income of the customer.
* **Educational Qualifications:** Educational qualifications of the customer.
* **Family Size:** Number of individuals in the customer's family.
* **Latitude:** Latitude of the customer's location.
* **Longitude:** Longitude of the customer's location.
* **Pin Code:** Pin code of the customer's location.
* **Output:** Status of the order.
* **Feedback:** Feedback provided by the customer after receiving the order.

## Purpose

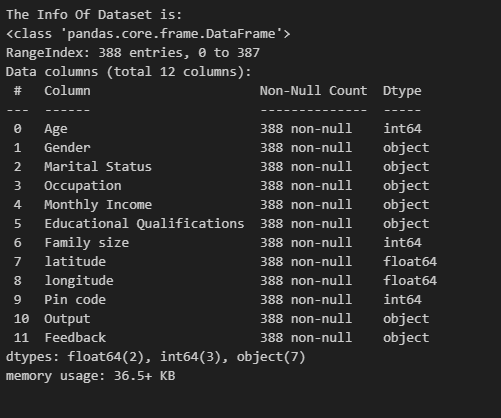
This dataset can be utilized to evaluating online food ordering platforms, integrating various user demographic and behaviour variables to assess and enhance user experience and platform performance. This tool aims to guide stakeholders through intense market competition by providing actionable insights and robust metrics for strategic decision-making and performance benchmarking.

## Head of Dataset

## Shape of Dataset



## Info of Dataset



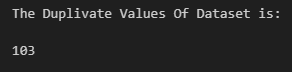
## Percent Missing of Dataset

## Null Values of Dataset

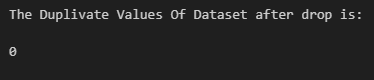
## Numerical Summary of Dataset



## Duplicate Values of Dataset

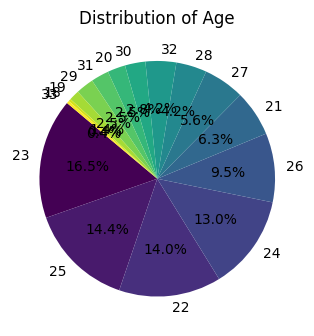
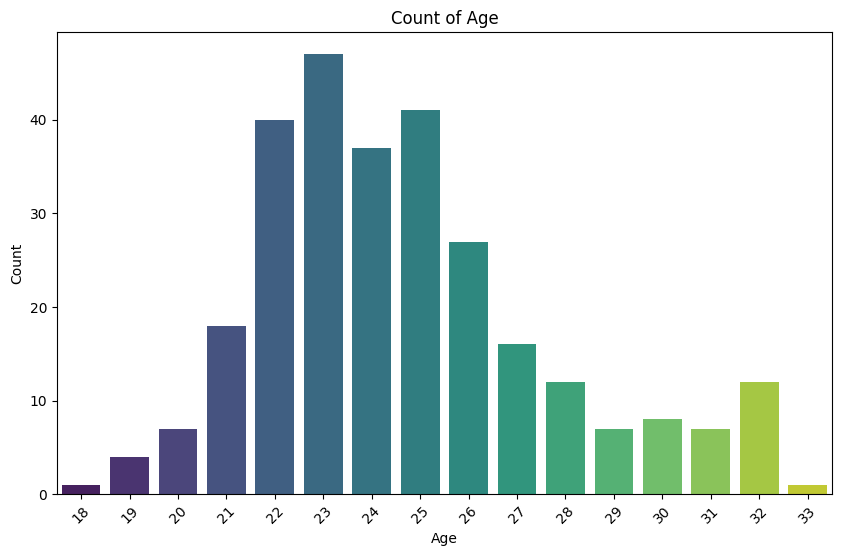


## Duplicate Values of Dataset after DROP

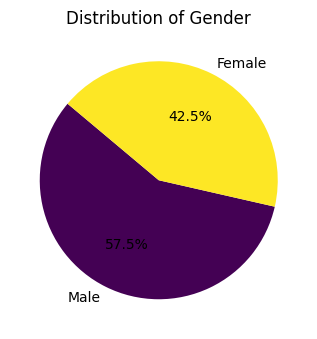
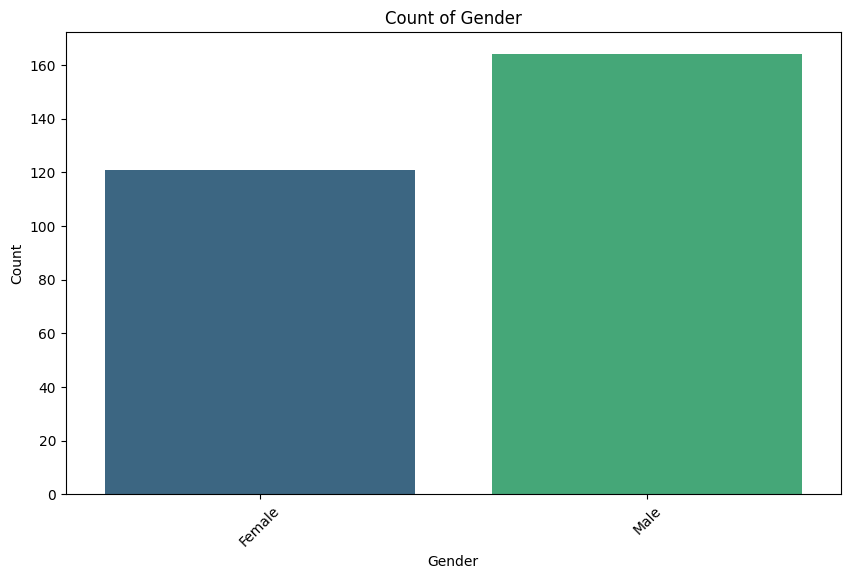


# Multivariate Analysis

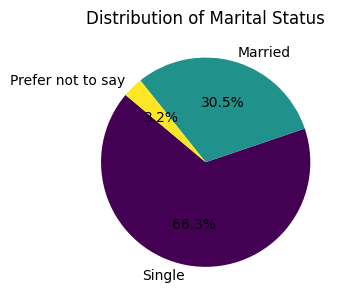
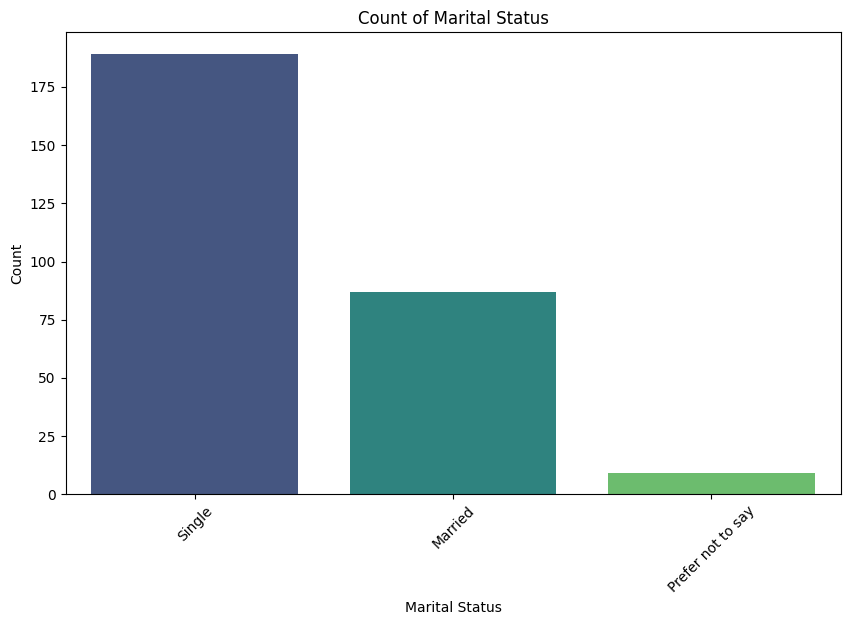
## Plot Distribution of Age



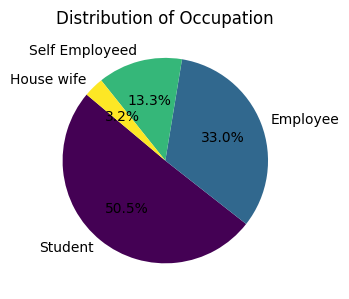
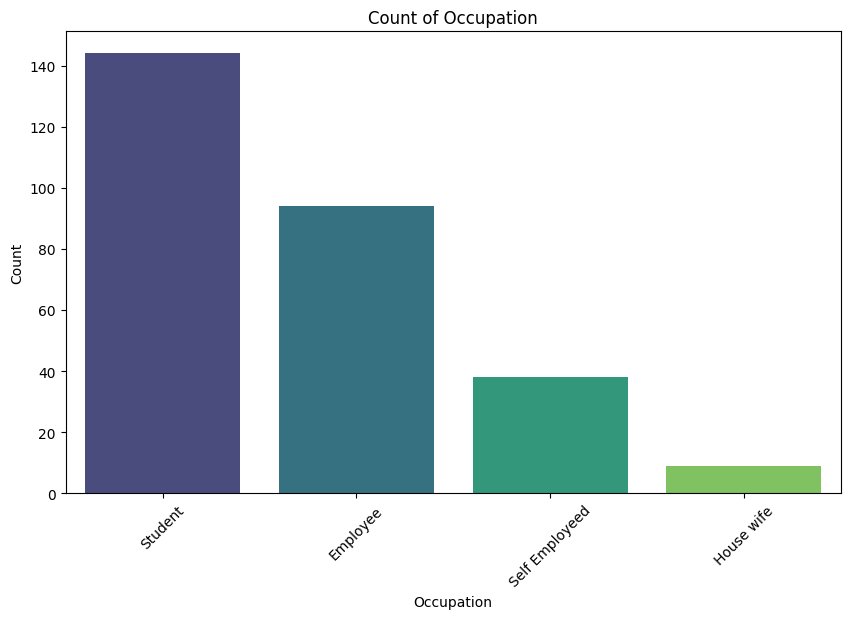
## Plot Distribution of Gender



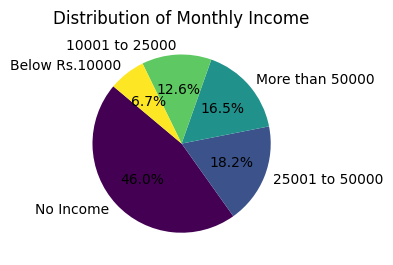
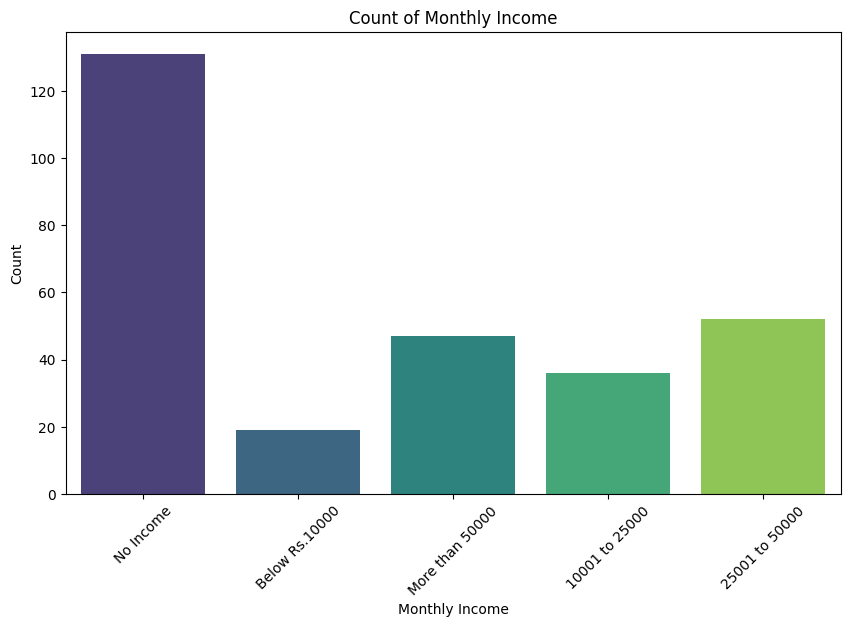
## Plot Distribution of Marital Status



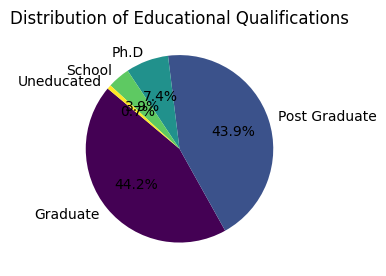
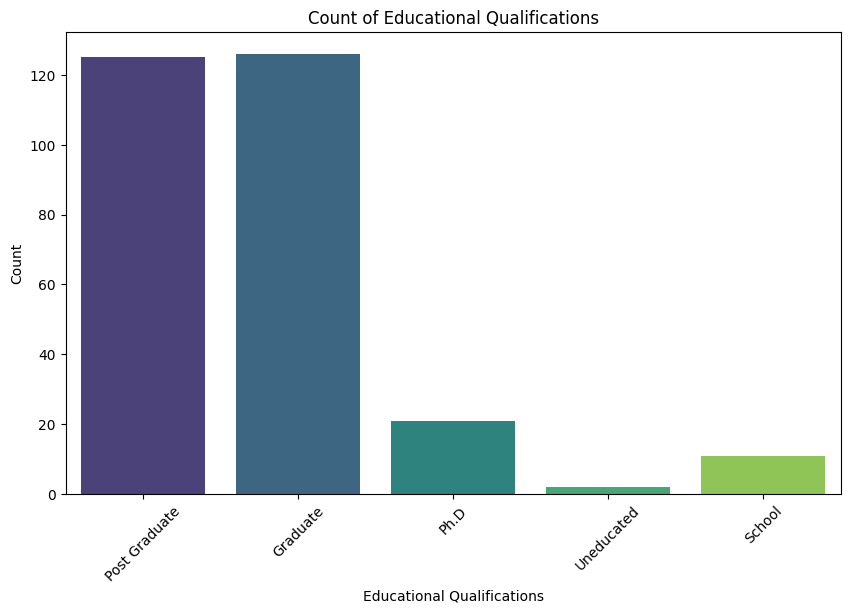
## Plot Distribution of Occupation



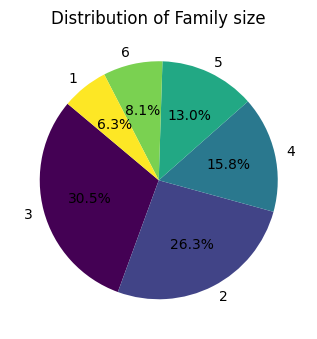
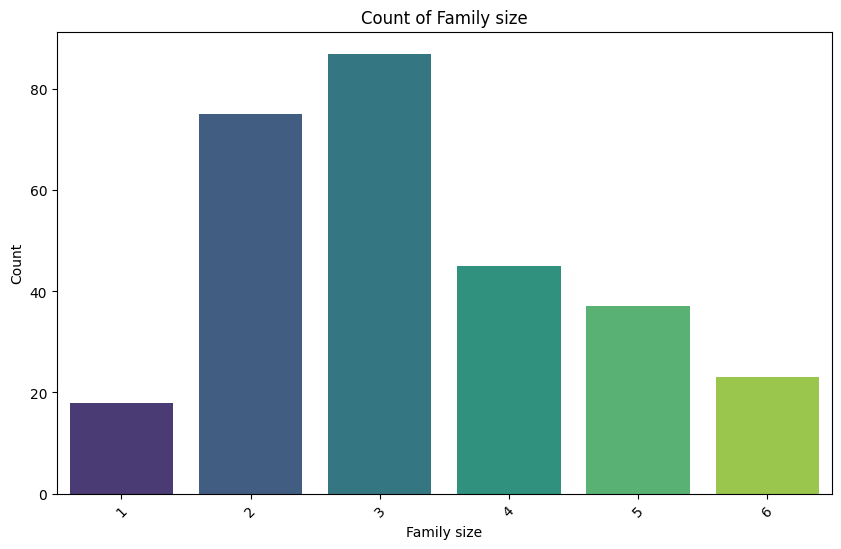
## Plot Distribution of Monthly Income



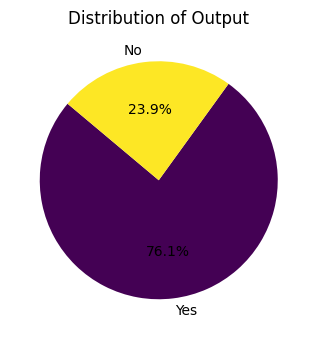
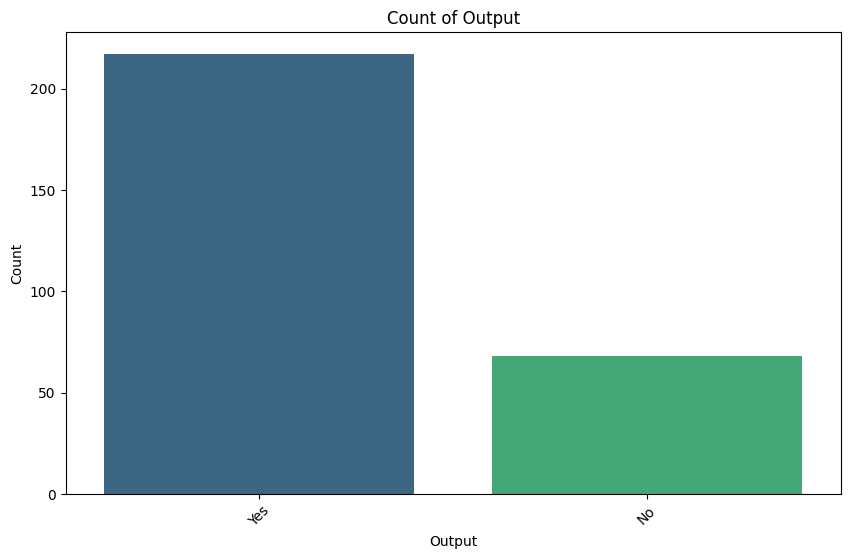
## Plot Distribution of Educational Qualifications



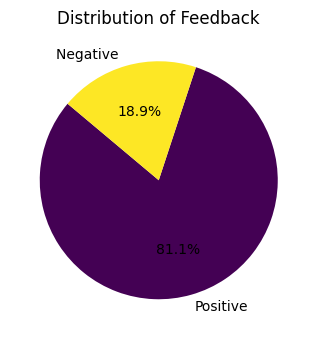
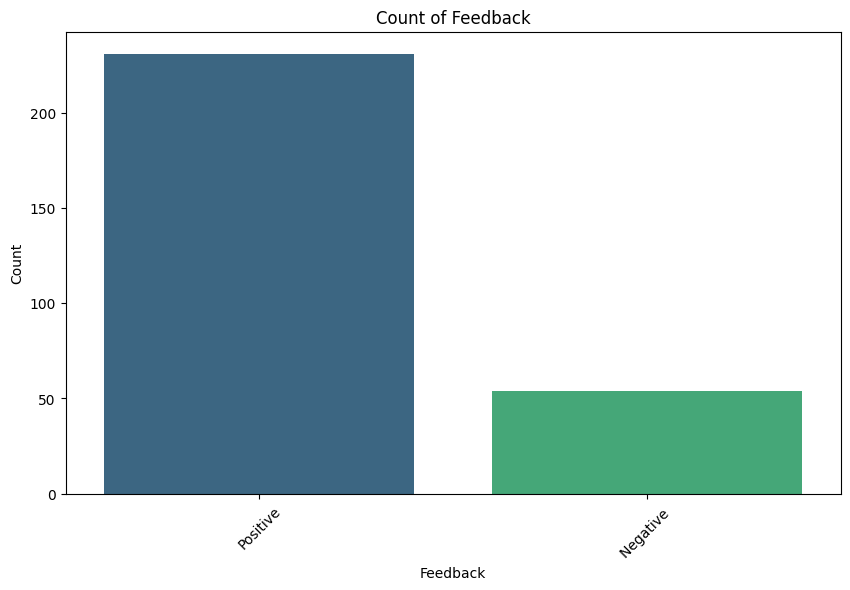
## Plot Distribution of Family Size



## Plot Distribution of Output



## Plot Distribution of Feedback



## PCA - Principal Component Analysis

### Normalization

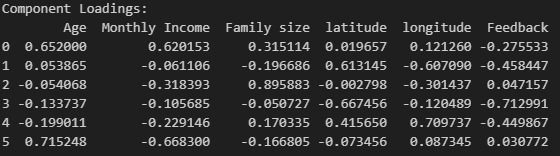
**variables** = ['Age', 'Monthly Income', 'Family size', 'latitude', 'longitude', 'Feedback']

### Explained Variance Ratio

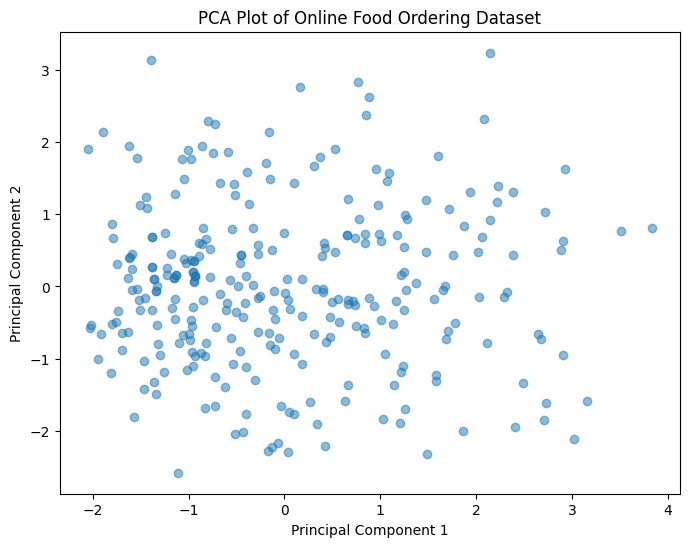


### Explained Variance Ratio by Principal Component

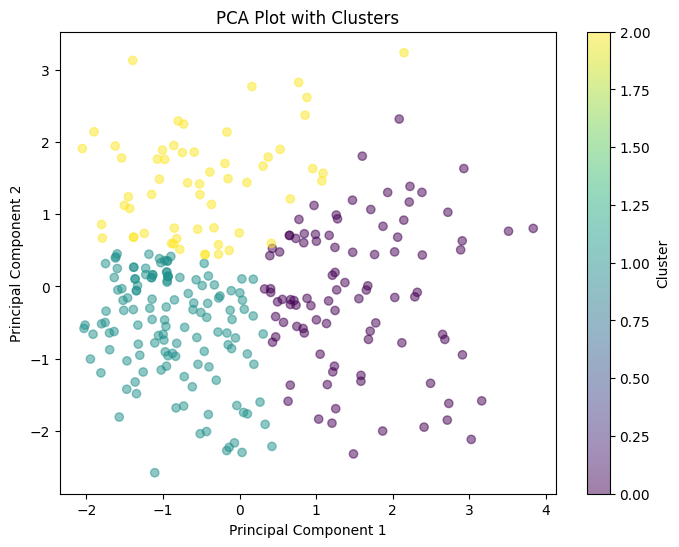
### Component Loading



### Plot of Online Food Ordering Dataset

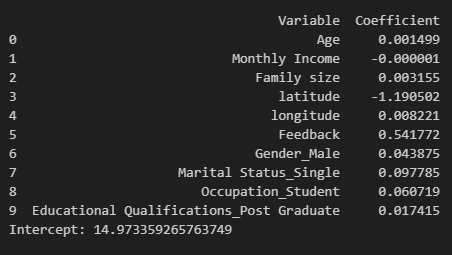


### Plot of Online Food Ordering Dataset with Clusters



## Multiple Regression

### Coefficient and Intercept



## PairPlot

### Normalization

**columns\_to\_plot** =['Age','Gender','Family size', 'latitude', 'longitude','Pin code',]

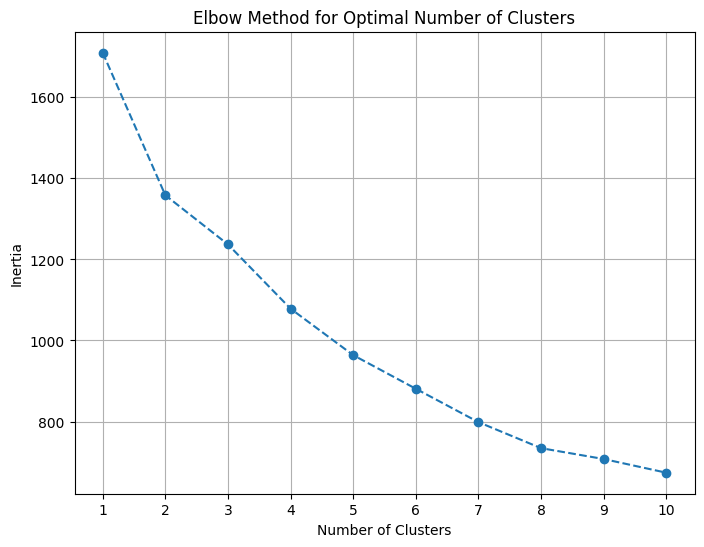


## Cluster Analysis

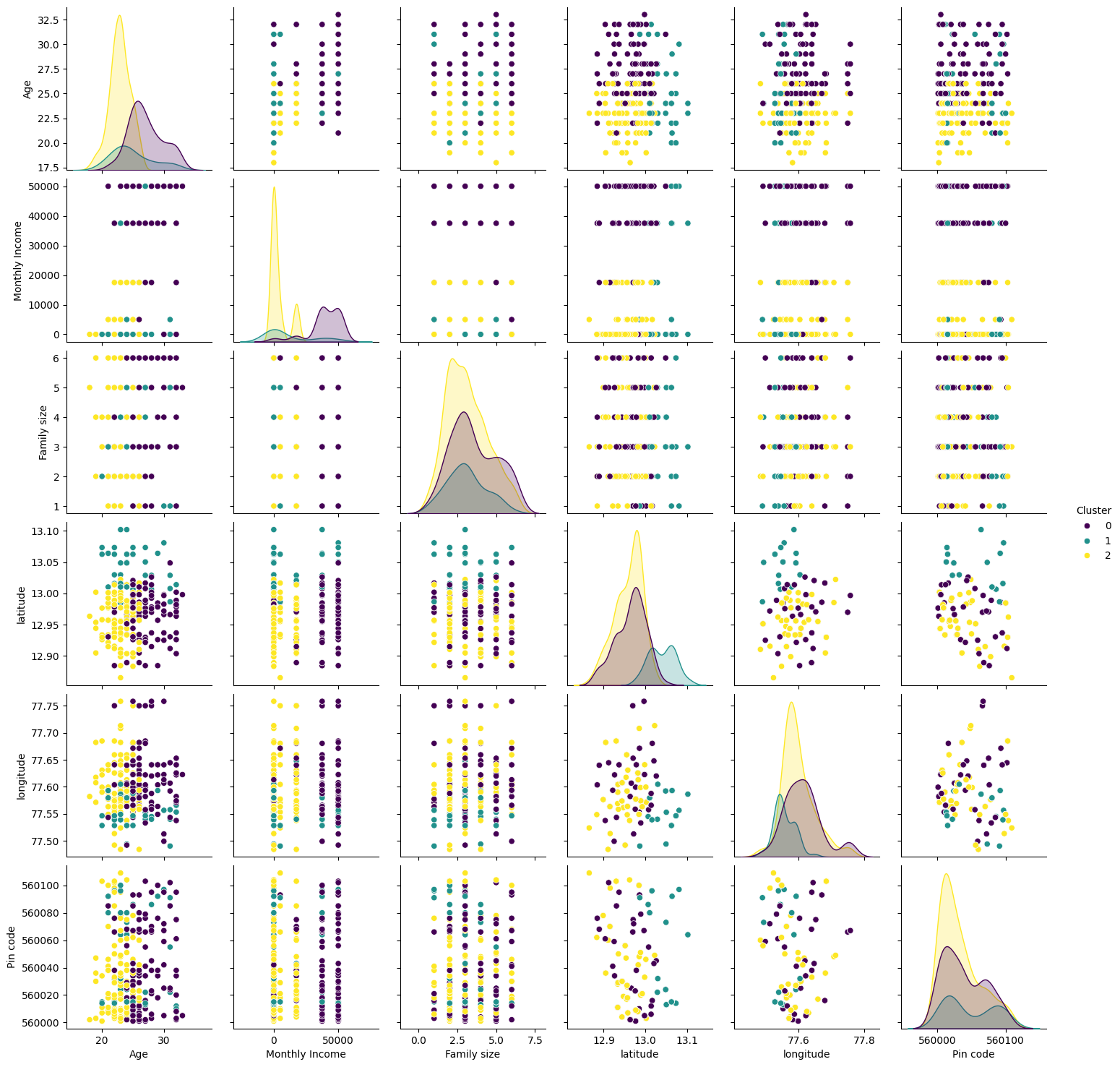
### Normalization

**features** = ['Age', 'Monthly Income', 'Family size', 'latitude', 'longitude', 'Pin code']

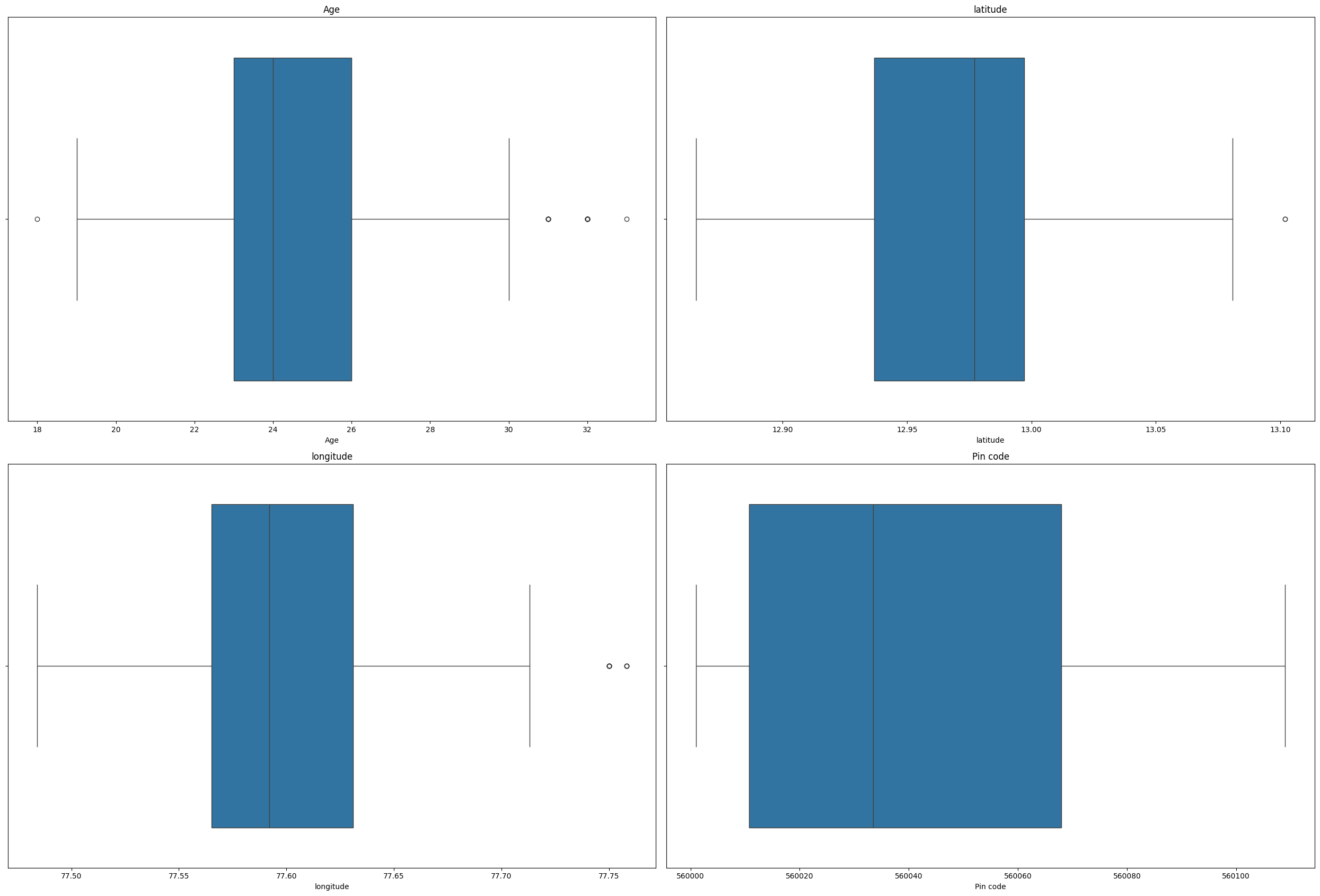
### Elbow Method for Optimal Number of Clusters



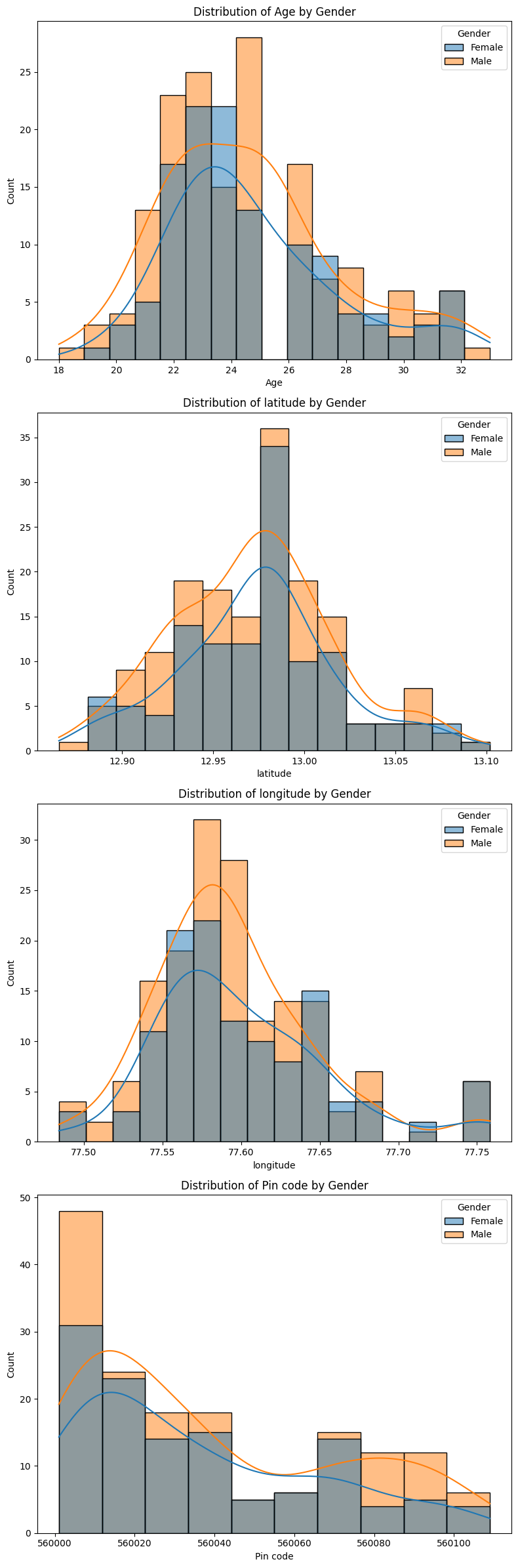
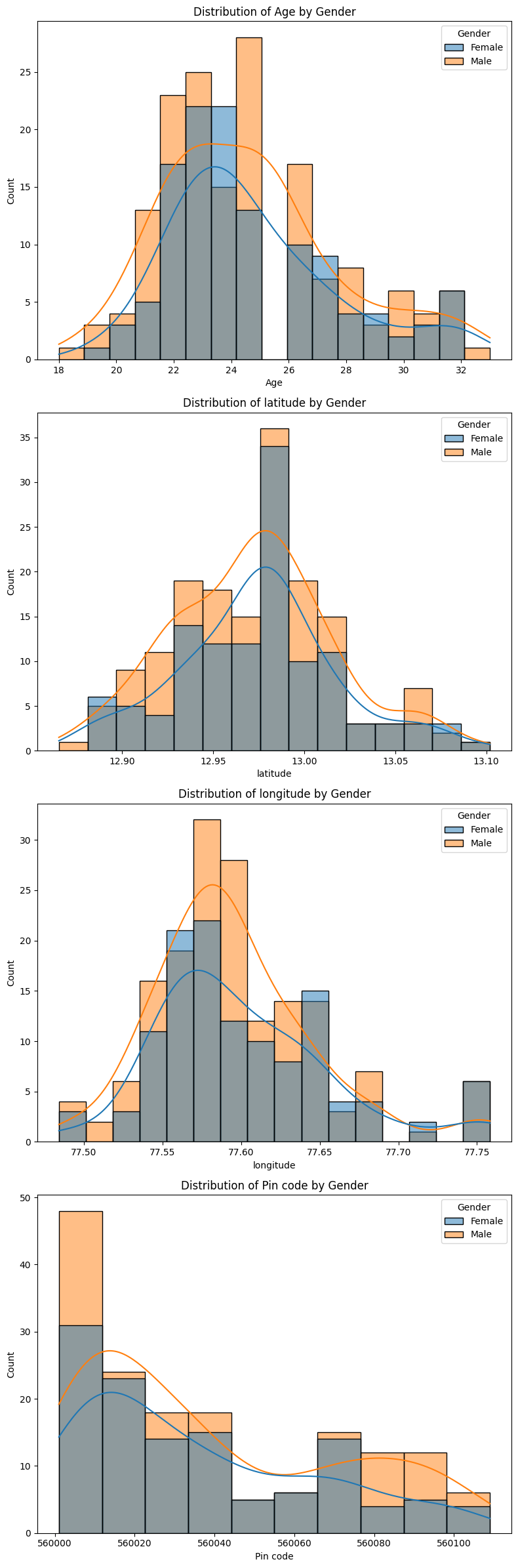
### Clusters PairPlot



## Outlier Detection



## Numerical Feature Distribution by Gender

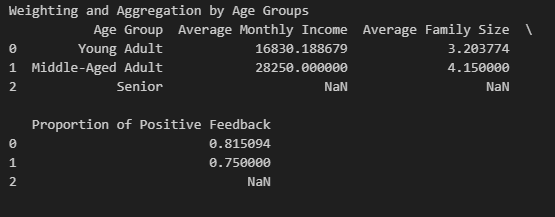


# Weighting and Aggregation

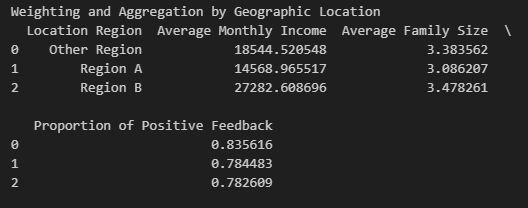
## Weighting and Aggregation by Gender

## 

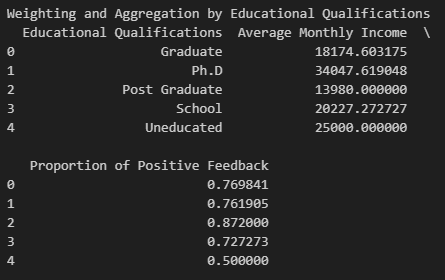
## Weighting and Aggregation by Age Groups



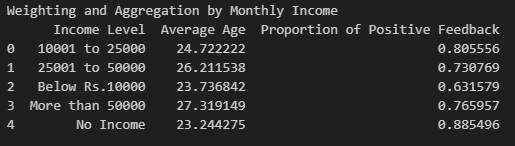
## Weighting and Aggregation by Geographic Location



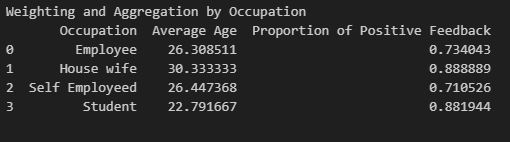
## Weighting and Aggregation by Educational Qualifications



## Weighting and Aggregation by Monthly Income



## Weighting and Aggregation by Occupation



# Link to other indices

## Hypothetical Composite Index: Online Food Ordering Experience Index (OFOEI)

To measure and track the user experience and performance of an online food ordering platform, focusing on customer satisfaction, delivery efficiency, and digital interface usability.

## Comparing OFOEI with Established Indices

### Comparison with Net Promoter Score (NPS) for Service Industries

**Benchmarking:**

* **Service Industry Average NPS:** 40
* **OFOEI NPS:** 55

**Analysis:**

The OFOEI’s higher NPS indicates stronger customer loyalty compared to the broader service industry, likely reflecting effective order fulfilment and exceptional customer service.

### Comparison with E-commerce User Experience Indexes

**Trend Analysis:**

* **E-commerce UX Index**: Emphasizes trends in mobile commerce usability.
* **OFOEI**: Shows significant improvements in mobile app navigation and ordering processes based on iterative user feedback.

**Insights:**

Both indices demonstrate improvements in mobile user experience, but OFOEI’s focused enhancements suggest a better alignment with consumer usage patterns, likely contributing to higher user satisfaction scores.