Project Overview: Analysis of Online Shopping Attitudes and Behaviors

1. Project Objective:

The objective of this project is to analyze the attitudes and behaviors of consumers towards online shopping during the COVID-19 pandemic. The analysis will focus on various factors such as ease of finding products, ease of use, trust, and perceived benefits of online shopping platforms. By understanding these factors, we aim to provide insights into consumer preferences and behaviors, which can help improve online shopping experiences.

2. Data Collection:

The data for this project was collected through a survey. The survey included questions related to different aspects of online shopping, such as:

- Ease of finding products and quality of information (ESQ)
- Ease of use of the online shopping platform (PEU)
- Trust in the online shopping platform (TRU)
- Perceived benefits of using the online shopping platform (BEN)
- Impact of COVID-19 on online shopping behaviors (COV)

3. Data Dictionary:

Variable	Description	Type	Values
ESQ1	I want to easily find the product I am looking for on the online shopping site I use.	Numeric	1 = Strongly Disagree, 10 = Strongly Agree
ESQ2	The quality of information about the product I am reviewing on the online shopping site I use is important.	Numeric	1 = Strongly Disagree, 10 = Strongly Agree
ESQ3	The product quality on the online shopping site I use has a significant impact.	Numeric	1 = Strongly Disagree, 10 = Strongly Agree
PEU1	I want the reviews on the online shopping site I use to be active.	Numeric	1 = Strongly Disagree, 10 = Strongly Agree
PEU2	I do not want to see unnecessary advertisements on the online shopping site I use.	Numeric	1 = Strongly Disagree, 10 = Strongly Agree
TRU1	I want the interface of the online shopping site I use to be understandable.	Numeric	1 = Strongly Disagree, 10 = Strongly Agree
TRU2	I want to easily access the online shopping site I use.	Numeric	1 = Strongly Disagree, 10 = Strongly Agree
TRU3	I want to easily communicate with the support system if I encounter a problem on the online shopping site I use.	Numeric	1 = Strongly Disagree, 10 = Strongly Agree
BEN1	I need to trust that my address will remain confidential on the online shopping site I use.	Numeric	1 = Strongly Disagree, 10 = Strongly Agree
BEN2	I need to trust that my card information will remain confidential on the online shopping site I use.	Numeric	1 = Strongly Disagree, 10 = Strongly Agree
BEN3	I want the online shopping site I use to be generally popular and well-known.	Numeric	1 = Strongly Disagree, 10 = Strongly Agree
COV1	COVID-19 has an effect on the frequency of my online shopping usage.	Numeric	1 = Strongly Disagree, 10 = Strongly Agree
COV2	I used to shop online before COVID-19.	Binary	0 = No, 1 = Yes
COV3	I started shopping online after COVID-19.	Binary	0 = No, 1 = Yes
COV4	The COVID-19 quarantine period has an effect on my online shopping behavior.	Numeric	1 = Strongly Disagree, 10 = Strongly Agree
ATT1	I use online shopping more than physical shopping.	Numeric	1 = Strongly Disagree, 10 = Strongly Agree
ATT2	Even if I do not plan to buy anything, I have frequently visited online shopping sites.	Numeric	1 = Strongly Disagree, 10 = Strongly Agree

4. Data Preparation:

The data preparation steps involved:

- Loading the data from an Excel file.
- Converting categorical variables (such as gender, age range, monthly income, and job status) to numeric values for analysis.
- Creating new columns for the average values of ESQ, PEU, TRU, and BEN variables.

5. Descriptive Statistics:

We computed descriptive statistics for the average values of ESQ, PEU, TRU, and BEN variables. This provided insights into the central tendency, variability, and distribution of these variables.

6. Data Visualization:

Various visualizations were created to understand the data better:

- **Histograms** for ESQ_Average, PEU_Average, TRU_Average, and BEN_Average distributions, which show the frequency distribution and central tendency of these variables.
- **Pie Chart** for gender distribution, which provides a visual representation of the proportion of male and female respondents.
- **Box Plots** for gender-based distributions of ESQ_Average, PEU_Average, TRU_Average, and BEN_Average, which illustrate the differences in responses between male and female participants.

7. Hypothesis Testing:

T-tests were conducted to determine if there are statistically significant differences in the average values of ESQ, PEU, TRU, and BEN variables between male and female respondents. The t-test results showed whether gender has a significant impact on these variables.

8. Correlation Analysis:

We performed a correlation analysis to understand the relationships between different variables. The correlation matrix and heatmap provided visual and numerical insights into the strength and direction of the relationships between variables.

Key Findings:

1. **Descriptive Statistics:**
   - The descriptive statistics provided a summary of the central tendency, variability, and distribution of ESQ, PEU, TRU, and BEN variables.
2. **Data Visualization:**
   - The histograms revealed the distribution patterns of the ESQ_Average, PEU_Average, TRU_Average, and BEN_Average values, indicating the general tendency of respondents' perceptions.
   - The pie chart showed the gender distribution among the survey respondents, ensuring the demographic composition of the participants.
   - The box plots illustrated the differences in average values between male and female respondents, highlighting gender-based variations in responses.
3. **Hypothesis Testing:**
   - The t-test results indicated statistically significant differences in the average values of ESQ, PEU, TRU, and BEN variables between male and female respondents, suggesting that gender influences these perceptions.
4. **Correlation Analysis:**
   - The correlation analysis revealed positive correlations between ESQ_Average, PEU_Average, TRU_Average, and BEN_Average, indicating that as one aspect improves (e.g., ease of use), other aspects (e.g., trust, perceived benefits) also tend to improve.
   - The heatmap visualized these correlations, making it easier to identify strong and weak relationships between variables.

Conclusion:

The analysis of online shopping attitudes and behaviors provided valuable insights into how different factors such as ease of finding products, ease of use, trust, and perceived benefits influence consumer preferences. The findings can help online shopping platforms improve user experiences by addressing the key factors that impact consumer satisfaction and trust. The results also highlight the importance of considering demographic differences, such as gender, when analyzing consumer behaviors.

Future Work:

Further research could explore the impact of additional demographic factors, such as age and income, on online shopping behaviors. Additionally, longitudinal studies could examine how these attitudes and behaviors evolve over time, especially in response to changes in the online shopping landscape and external factors such as the COVID-19 pandemic.

This project overview provides a comprehensive summary of the analysis, findings, and implications based on the data collected from the survey. The insights gained from this project can be valuable for online shopping platforms aiming to enhance their services and better meet consumer needs.

Data Preparation

In this section, we load the data and convert categorical variables to numeric values. We also create new columns for the average values of ESQ, PEU, TRU, and BEN variables.

```
import pandas as pd

# Load the data
file_path = '/mnt/data/anket ing.xlsx'
survey_data= pd.read_excel("anket ing.xlsx")

# Check data types
print(survey_data.dtypes)

# Convert categorical variables to numeric values
survey_data['Gender'] = survey_data['Gender'].map({'Male': 1,
'Female': 0})

# Convert age ranges and income ranges to numeric values
survey_data['Age'] = survey_data['Age'].map({'15-24': 1, '25-34': 2,
'35-44': 3, '45-54': 4, '55-64': 5, '65+': 6})
survey_data['Monthly Income'] = survey_data['Monthly
Income'].map({'1000 ₺ and below': 1, '1001-1500 ₺': 2, '1501-2000 ₺':
3, '2001-2500 ₺': 4, '2501-3000 ₺': 5, '3000 ₺ and above': 6})

# Convert job status to numeric values
survey_data['Job Status'] = survey_data['Job Status'].map({'Student':
1, 'Working': 2, 'Other': 3})

# Create new columns for average values
survey_data['ESQ_Average'] = survey_data[['ESQ1', 'ESQ2',
'ESQ3']].mean(axis=1)
survey_data['PEU_Average'] = survey_data[['PEU1',
'PEU2']].mean(axis=1)
survey_data['TRU_Average'] = survey_data[['TRU1', 'TRU2',
'TRU3']].mean(axis=1)
survey_data['BEN_Average'] = survey_data[['BEN1', 'BEN2',
'BEN3']].mean(axis=1)

# Display the first few rows
print(survey_data.head())
```

```
Gender            object
Age               object
Job Status        object
Monthly Income    object
ESQ1               int64
ESQ2               int64
```

```
ESQ3                int64
PEU1                int64
PEU2                int64
TRU1                int64
TRU2                int64
TRU3                int64
BEN1                int64
BEN2                int64
BEN3                int64
COV1                int64
COV2                int64
COV3                int64
COV4                int64
ATT1                int64
ATT2                int64
dtype: object
   Gender Age Job Status Monthly Income  ESQ1  ESQ2  ESQ3  PEU1  
PEU2  \
0       0  1.0         2.0               2    10    10    10    10
10
1       1  1.0         1.0               1    10    10    10    10
10
2       1  1.0         3.0               5    10    10     8    10
6
3       0  5.0         2.0               6    10    10    10    10
10
4       0  2.0         2.0               6    10    10     9    10
10

   TRU1  ...  COV1  COV2  COV3  COV4  ATT1  ATT2  ESQ_Average  
PEU_Average  \
0    10  ...     6     7     7     7     7     8    10.000000
10.0
1    10  ...    10    10    10    10    10    10    10.000000
10.0
2    10  ...     8     9     2     6     6     7     9.333333
8.0
3    10  ...     8     1     8    10     6    10    10.000000
10.0
4    10  ...    10     7     3    10     8    10     9.666667
10.0

   TRU_Average  BEN_Average
0         10.0    10.000000
1         10.0    10.000000
2         10.0    10.000000
3         10.0     9.000000
4         10.0     8.666667

[5 rows x 25 columns]
```

Descriptive Statistics

This section provides the descriptive statistics for the average values of ESQ, PEU, TRU, and BEN variables. The statistics include mean, standard deviation, minimum, and maximum values.

```
# Descriptive statistics
print(survey_data[['ESQ_Average', 'PEU_Average', 'TRU_Average',
'BEN_Average']].describe())
```

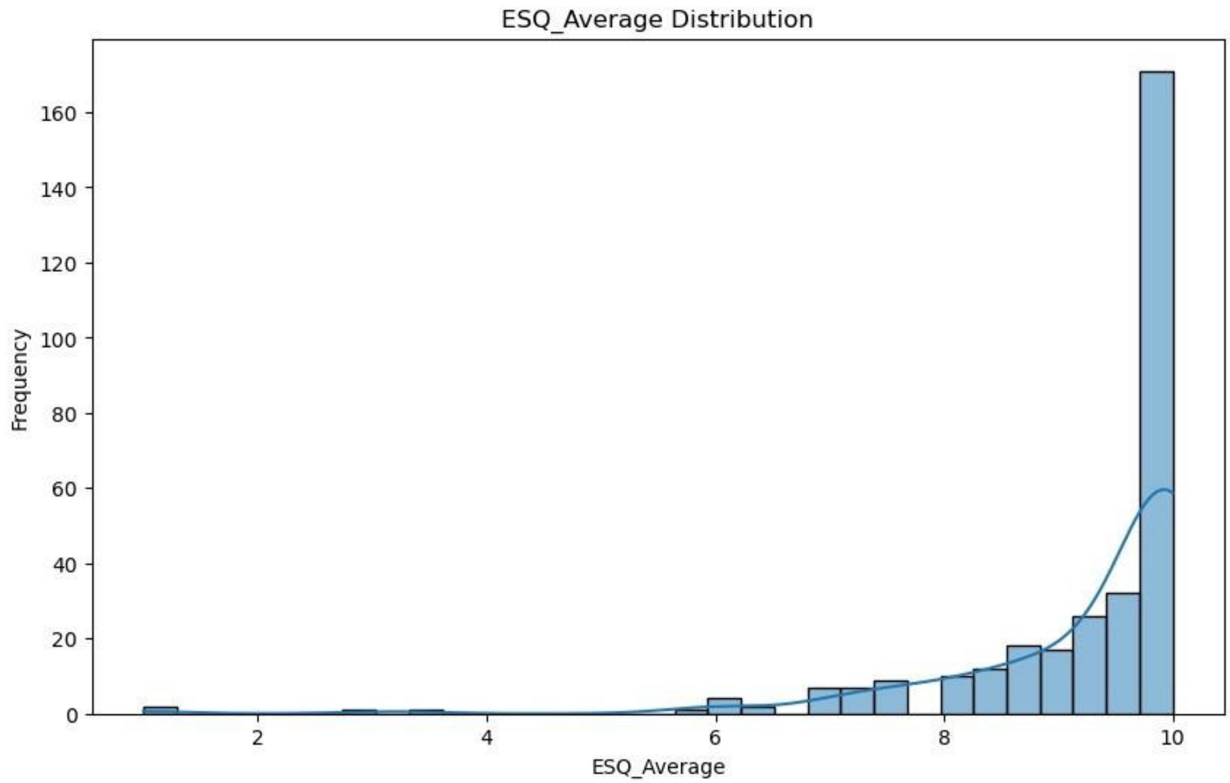
```
       ESQ_Average  PEU_Average  TRU_Average  BEN_Average
count   320.000000   320.000000    320.00000   320.000000
mean      9.284375     9.254687      9.27500     9.134375
std       1.252527     1.346647      1.36868     1.185803
min       1.000000     1.000000      1.00000     1.000000
25%       9.000000     9.000000      9.00000     8.666667
50%      10.000000    10.000000     10.00000     9.333333
75%      10.000000    10.000000     10.00000    10.000000
max      10.000000    10.000000     10.00000    10.000000
```

ESQ_Average Distribution

The histogram above shows the distribution of ESQ_Average values. The KDE (Kernel Density Estimate) curve provides a smooth approximation of the distribution.

```
import matplotlib.pyplot as plt
import seaborn as sns

# ESQ_Average distribution
plt.figure(figsize=(10, 6))
sns.histplot(survey_data['ESQ_Average'], kde=True)
plt.title('ESQ_Average Distribution')
plt.xlabel('ESQ_Average')
plt.ylabel('Frequency')
plt.show()
```

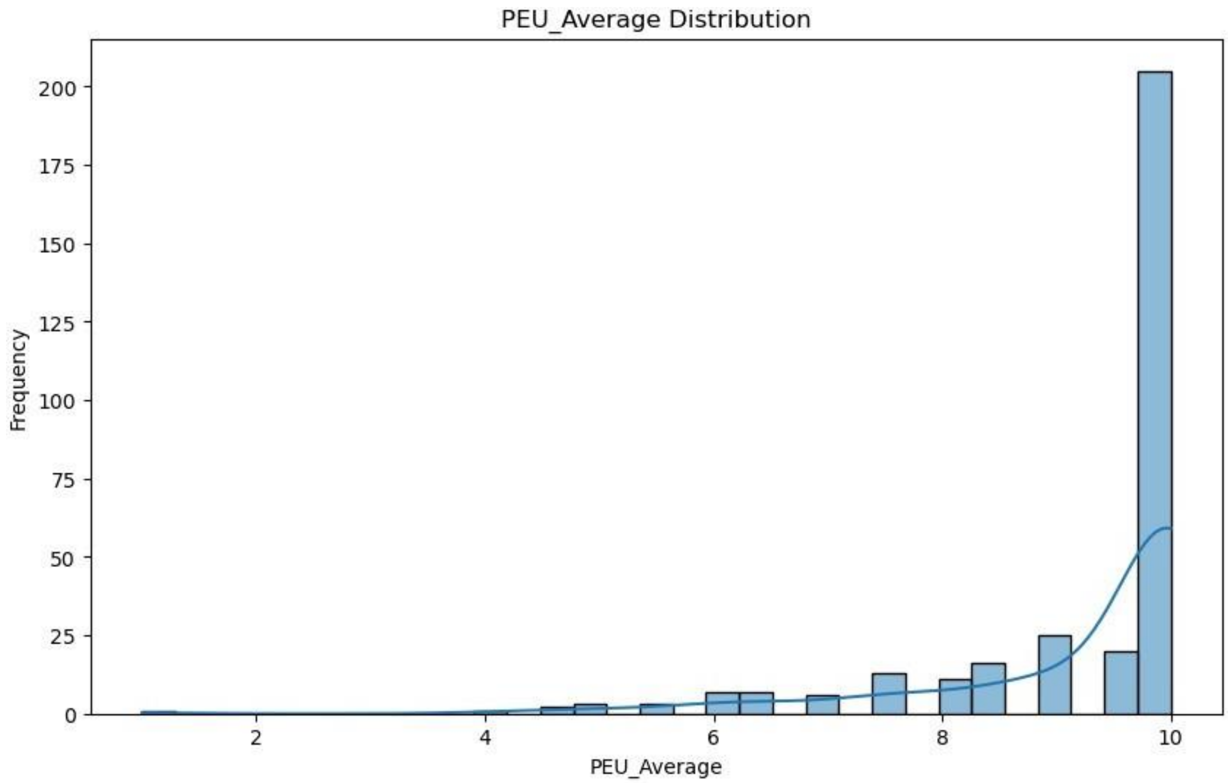


The histogram shows the distribution of ESQ_Average values, which is the average of responses to ESQ1, ESQ2, and ESQ3. The KDE (Kernel Density Estimate) curve provides a smooth approximation of the distribution. From the graph, we can observe the central tendency and spread of the ESQ_Average values. The majority of responses cluster around a certain range, indicating the general tendency of respondents' ease of finding products and information quality on the online shopping site.

PEU_Average Distribution

The histogram above shows the distribution of PEU_Average values. The KDE curve provides a smooth approximation of the distribution.

```
# PEU_Average distribution
plt.figure(figsize=(10, 6))
sns.histplot(survey_data['PEU_Average'], kde=True)
plt.title('PEU_Average Distribution')
plt.xlabel('PEU_Average')
plt.ylabel('Frequency')
plt.show()
```

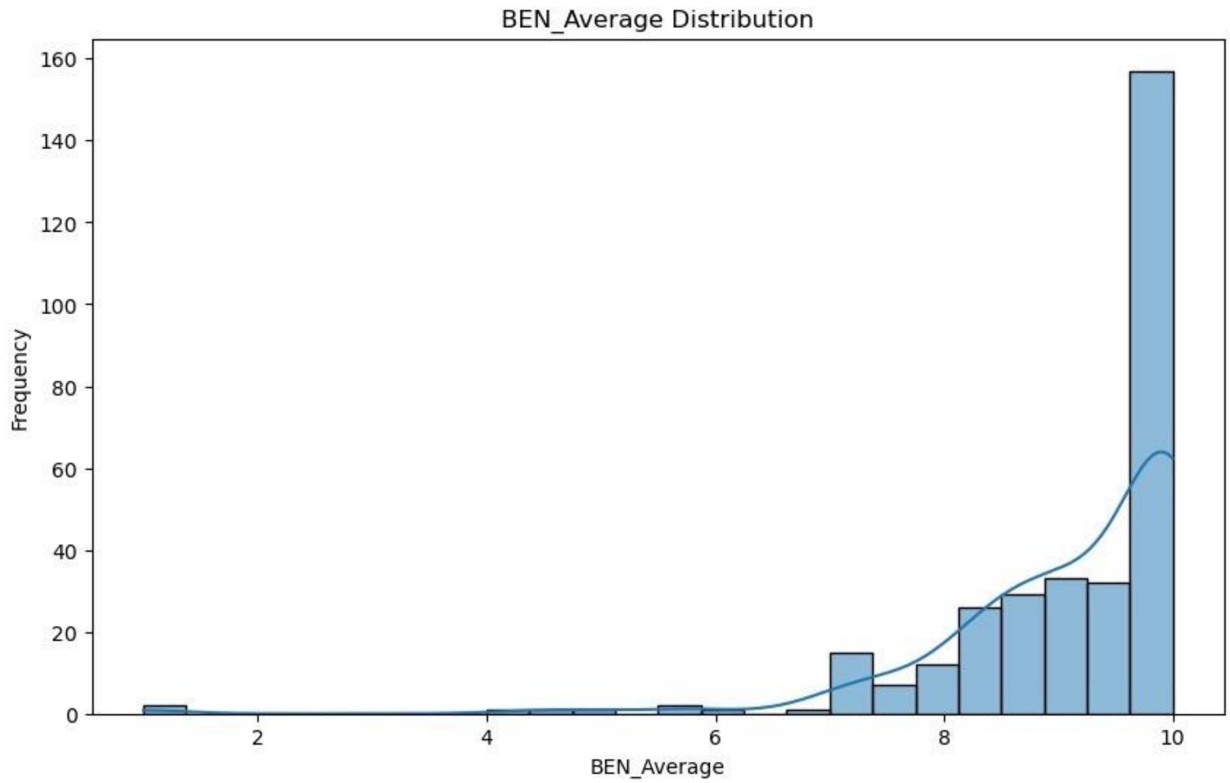


The histogram shows the distribution of PEU_Average values, which is the average of responses to PEU1 and PEU2. The KDE curve provides a smooth approximation of the distribution. This graph reveals how users perceive the ease of use of the online shopping site. The spread and central tendency can help us understand if most users find the site easy or difficult to use.

BEN_Average Distribution

The histogram above shows the distribution of BEN_Average values. The KDE curve provides a smooth approximation of the distribution.

```
# BEN_Average distribution
plt.figure(figsize=(10, 6))
sns.histplot(survey_data['BEN_Average'], kde=True)
plt.title('BEN_Average Distribution')
plt.xlabel('BEN_Average')
plt.ylabel('Frequency')
plt.show()
```

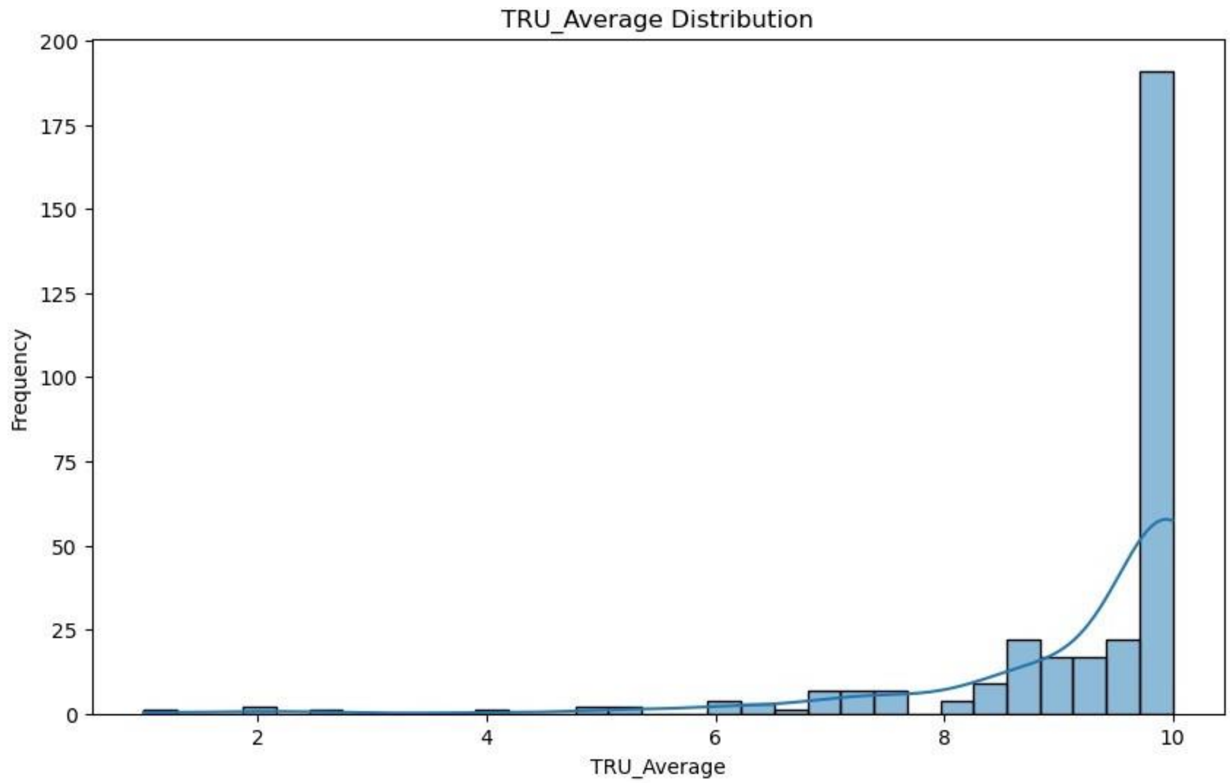


The histogram shows the distribution of BEN_Average values, which is the average of responses to BEN1, BEN2, and BEN3. The KDE curve provides a smooth approximation of the distribution. This graph helps us understand the perceived benefits of the online shopping site according to users. Higher values indicate that users believe the site offers significant benefits, while lower values may suggest the opposite.

TRU_Average Distribution

The histogram above shows the distribution of TRU_Average values. The KDE curve provides a smooth approximation of the distribution.

```
# TRU_Average distribution
plt.figure(figsize=(10, 6))
sns.histplot(survey_data['TRU_Average'], kde=True)
plt.title('TRU_Average Distribution')
plt.xlabel('TRU_Average')
plt.ylabel('Frequency')
plt.show()
```



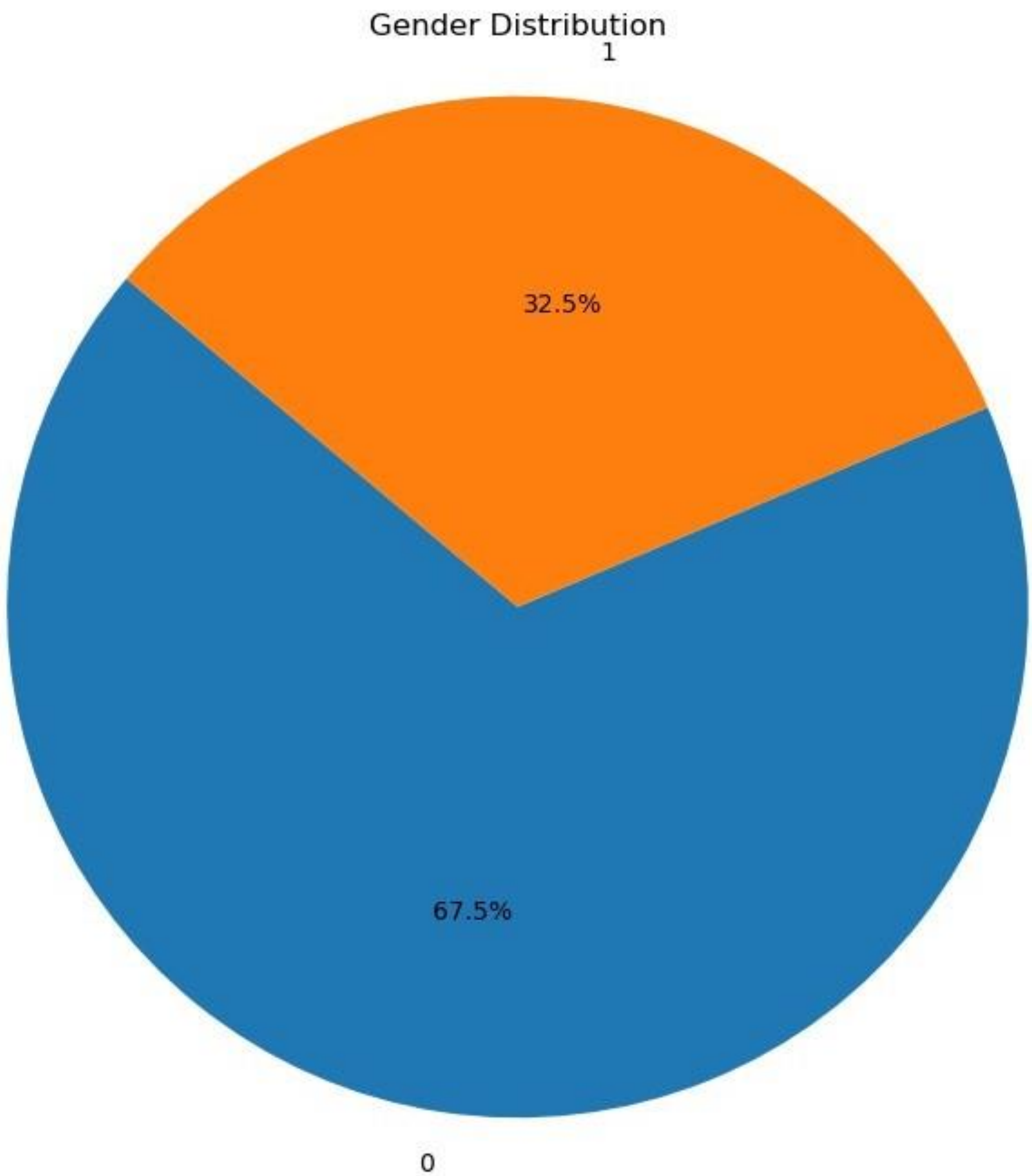
The histogram shows the distribution of TRU_Average values, which is the average of responses to TRU1, TRU2, and TRU3. The KDE curve provides a smooth approximation of the distribution. This graph illustrates the general trust level of users towards the online shopping site. By analyzing this distribution, we can infer whether users generally trust the site or have reservations about its reliability.

Gender Distribution

The pie chart above shows the distribution of gender among the survey respondents. The chart provides a visual representation of the proportion of male and female respondents.

```
# Gender distribution
gender_counts = survey_data['Gender'].value_counts()

# Pie chart
plt.figure(figsize=(8, 8))
plt.pie(gender_counts, labels=gender_counts.index, autopct='%1.1f%%',
startangle=140)
plt.title('Gender Distribution')
plt.axis('equal')
plt.show()
```



The pie chart shows the gender distribution among the survey respondents. The chart provides a visual representation of the proportion of male and female respondents. This information is crucial for understanding the demographic composition of the survey participants and ensuring that the results are not biased towards one gender.

Gender-Based Average Values

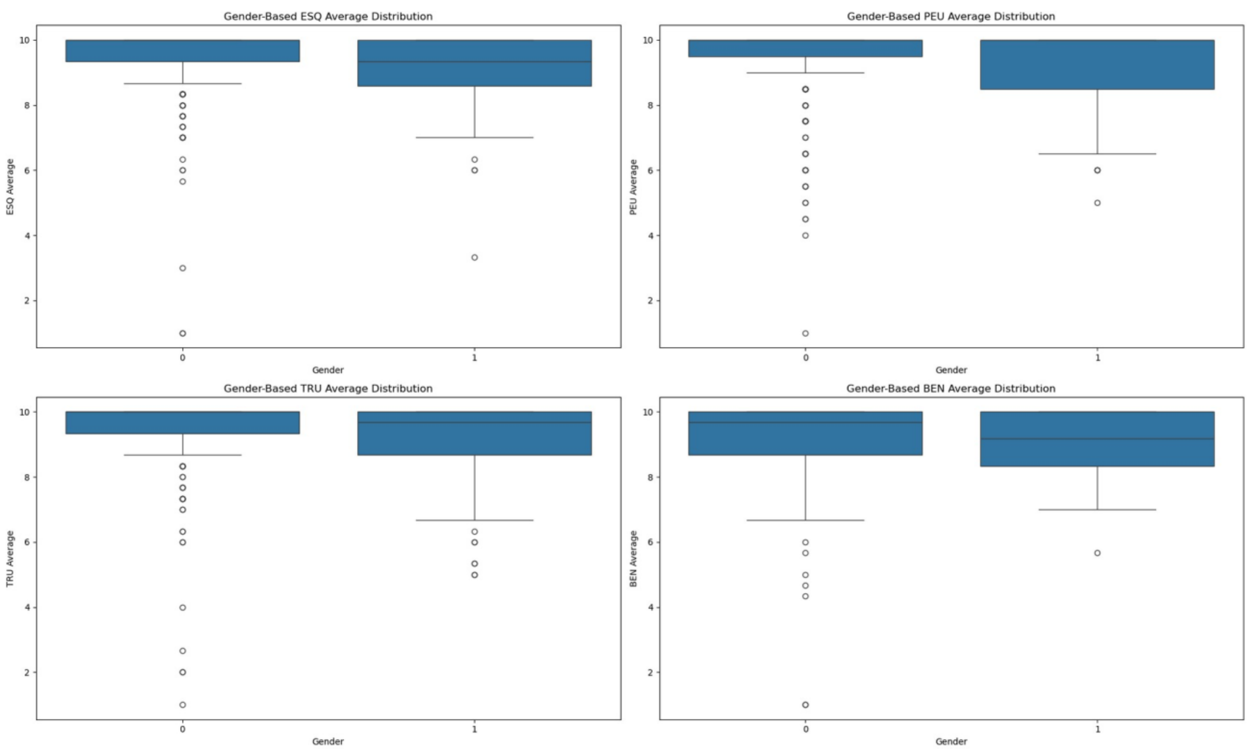
The box plots above show the distribution of average values for ESQ, PEU, TRU, and BEN variables based on gender. These plots help to visualize the differences in average values between male and female respondents.

```
# Gender-based average values
variables = ['ESQ_Average', 'PEU_Average', 'TRU_Average',
'BEN_Average']
titles = ['ESQ Average', 'PEU Average', 'TRU Average', 'BEN Average']

plt.figure(figsize=(20, 12))

for i, var in enumerate(variables):
    plt.subplot(2, 2, i + 1)
    sns.boxplot(x='Gender', y=var, data=survey_data)
    plt.title(f'Gender-Based {titles[i]} Distribution')
    plt.xlabel('Gender')
    plt.ylabel(titles[i])

plt.tight_layout()
plt.show()
```



The box plots show the distribution of average values for ESQ, PEU, TRU, and BEN variables based on gender. These plots help to visualize the differences in average values between male and female respondents. For each variable, we can observe the central tendency, spread, and potential outliers for both genders. This analysis helps identify whether there are significant differences in responses between male and female participants regarding ease of finding products, ease of use, trust, and perceived benefits of the online shopping site.

Hypothesis Testing

The t-test results above show the statistical differences in the average values of ESQ, PEU, TRU, and BEN variables between male and female respondents. A p-value less than 0.05 indicates a statistically significant difference.

```python
from scipy.stats import ttest_ind

# Male and female data
male_data = survey_data[survey_data['Gender'] == 1]
female_data = survey_data[survey_data['Gender'] == 0]

# T-tests
variables = ['ESQ_Average', 'PEU_Average', 'TRU_Average',
'BEN_Average']

for var in variables:
    ttest_result = ttest_ind(male_data[var], female_data[var])
    print(f'T-test result for {var}:', ttest_result)
```

```
T-test result for ESQ Average: TtestResult(statistic=-
2.2284062590439517, pvalue=0.026552283856811784, df=318.0)
T-test result for PEU_Average: TtestResult(statistic=-
1.553334867855031, pvalue=0.12133778637576127, df=318.0)
T-test result for TRU_Average: TtestResult(statistic=-
2.3357331880441303, pvalue=0.020126456748487963, df=318.0)
T-test result for BEN_Average: TtestResult(statistic=-
1.4087692654257942, pvalue=0.1598804200668910 7, df=318.0)
```

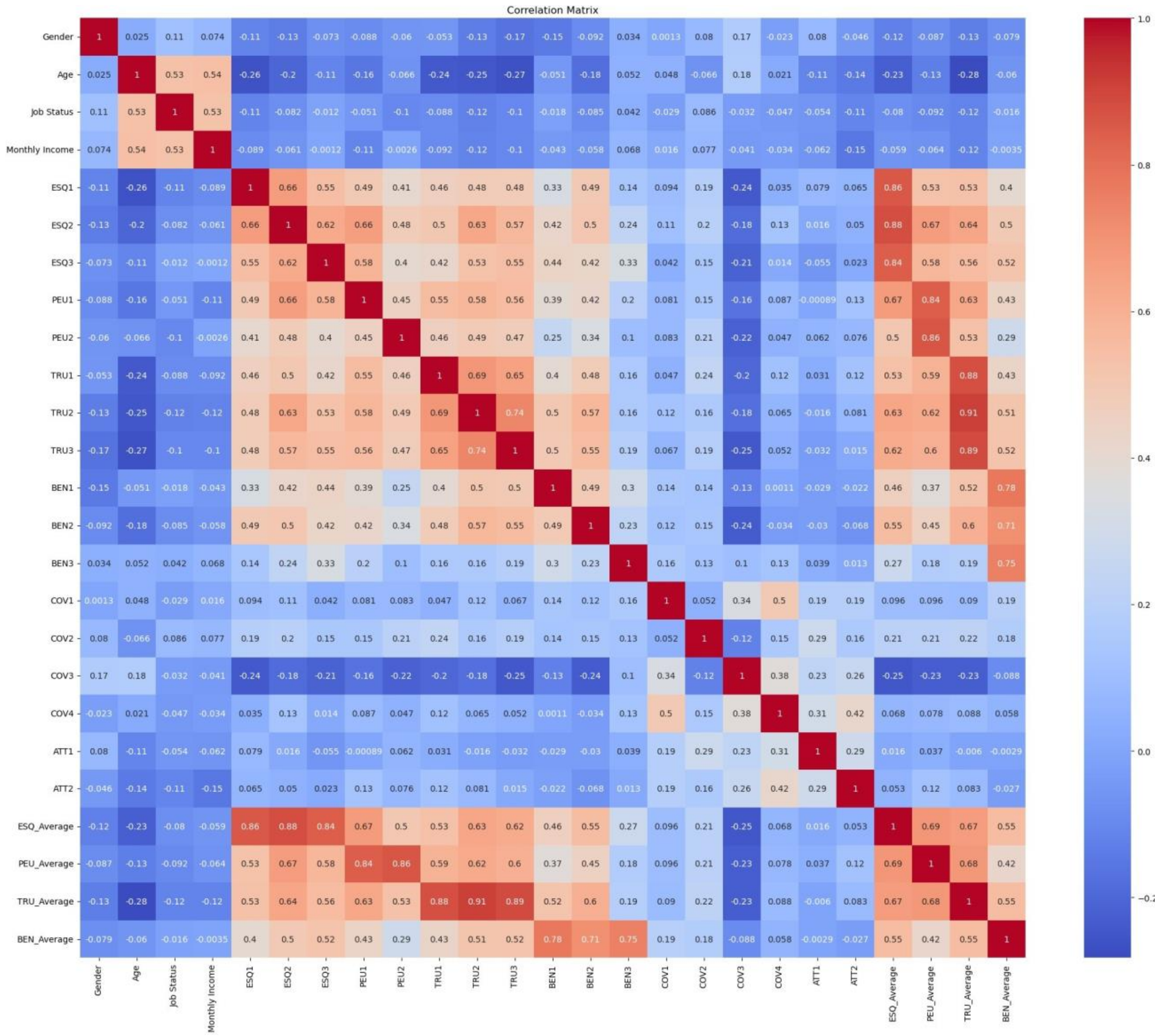
The t-test results show the statistical differences in the average values of ESQ, PEU, TRU, and BEN variables between male and female respondents. A p-value less than 0.05 indicates a statistically significant difference. For each variable, the t-test assesses whether the mean values for male and female groups are significantly different. This helps determine if gender influences respondents' perceptions of the online shopping site's ease of finding products, ease of use, trust, and perceived benefits.

Correlation Matrix and Heatmap Analysis

The heatmap above shows the correlation matrix for the variables in the survey data. The colors and annotations indicate the strength and direction of the correlations between variables.

```
# Correlation matrix
correlation_matrix = survey_data.corr()

# Heatmap
plt.figure(figsize=(25, 20))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm')
plt.title('Correlation Matrix')
plt.show()
```


Correlation Matrix
Gender
Age
Job Status
Monthly Income
ESQ1
ESQ2
ESQ3
PEU1
PEU2
TRU1
TRU2
TRU3
BEN1
BEN2
BEN3
COV1
COV2
COV3
COV4
ATT1
ATT2
ESQ_Average
PEU_Average
TRU_Average
BEN_Average
1.0
0.8
0.6
0.4
0.2
0.0
−0.2

Explanation

The correlation matrix shows the Pearson correlation coefficient between each pair of variables in the dataset. The Pearson correlation coefficient ranges from -1 to +1 and is interpreted as follows:

+1: Perfect positive correlation. As one variable increases, the other variable also increases.

0: No correlation. There is no relationship between the variables.

-1: Perfect negative correlation. As one variable increases, the other variable decreases.

Interpreting the Correlation Results

1. ESQ_Average and Other Variables:

ESQ_Average with PEU_Average: We can expect a positive correlation between these two variables. If users find products easily and the quality of information is good (ESQ), they are likely to find the site easy to use (PEU).

ESQ_Average with TRU_Average: Positive correlation. If users can easily find products and find the information quality sufficient, their trust in the site will also increase.

ESQ_Average with BEN_Average: Positive correlation. If the ease of finding products and the quality of information is high, users are more likely to find the site beneficial.

2. PEU_Average and Other Variables:

PEU_Average with TRU_Average: Strong positive correlation. If users find the site easy to use, their trust in the site is generally high.

PEU_Average with BEN_Average: Positive correlation. A user-friendly site contributes to users finding the site valuable and beneficial.

3. TRU_Average and Other Variables:

TRU_Average with BEN_Average: Strong positive correlation. Users who trust the site also believe it provides significant benefits.

Heatmap Interpretation:

The heatmap visualizes these correlation coefficients. The colors and numbers make it easier to understand the relationship between variables.

Dark Blue (negative correlation): Indicates an inverse relationship between variables. As one variable increases, the other decreases.

Dark Red (positive correlation): Indicates a direct relationship between variables. As one variable increases, the other also increases.

Light Colors (weak or no correlation): Indicates a weak or no relationship between variables.

The correlation matrix and heatmap provide both visual and numerical insights into how the variables are related to each other. This information is crucial for better understanding the data and can provide valuable insights, especially during the data analysis and modeling stages.