Analysis of Consumer Shopping Preferences

Ahmed Abdullah Shahid

Liverpool Hope University

COMM015AZ2024/5 - Data Analytics

24012028@hope.ac.uk

Table of Contents

1. Introduction:

Over the last decade, the way we shop has changed dramatically, thanks to the rise in e-commerce. Online shopping has become a favorite for many people, offering the convenience of browsing from home, access to a wide variety of products, and competitive prices. Plus, advancements in technology have made shopping online faster and easier than ever. However, despite these benefits, there is still a strong pull toward in-store shopping for some. Many people enjoy the instant fulfilment of picking up a product right away or the personal touch of attentive customer service, things that can be harder to come by when shopping online. So, while online shopping has taken over in many ways, there's still something special about the in-store experience that keeps it relevant.

For businesses looking to improve their strategies, it's essential to really understand what makes customers tick, what keeps them happy, how they like to shop, and how often they make purchases. This means diving deep into why people choose one shopping method over another and identifying key areas where they can enhance the overall shopping experience. Things like website quality, convenience, security and privacy, product quality, and logistics can all have a huge impact on how customers feel about their experience. When businesses pay attention to these details, they're more likely to keep customers satisfied and coming back for more.

Consumer behavior is always evolving, especially with the rapid growth of digital payment options, mobile shopping apps, and e-commerce platforms. As customers' shopping habits shift, it's important to understand how they feel about shopping online versus in-store, what influences their buying decisions, and what makes them happy with their shopping experience. By diving into these preferences and patterns, businesses can gain valuable insights that will help them tailor their products and services to better meet customer needs. Ultimately, this research will help companies create a more satisfying and personalized shopping experience for their customers.

2. Objectives:

The primary objectives of this study are:

- Compare online vs. in-store shopping preferences, understanding how many consumers prefer online shopping versus in-store shopping. Identify key reasons for their preferences.
- Analyze shopping frequency trends, determine how often consumers shop online and in-store. Identify patterns in shopping habits.
- Identifying key factors affecting customer satisfaction, evaluating the importance of website quality, security and privacy, product quality, logistics, and convenience in customer decisions. Measure customer perceptions of these factors.
- Understand how demographics influence shopping preferences, Analyze how age and gender impact shopping habits. Determine whether younger consumers prefer online shopping more than older generations.

3. Methodology:

To analyze consumer shopping preferences between online and in-store shopping, a data-driven approach was employed using Python for statistical analysis and visualization. The study follows these key steps:

3.1. Data Processing:

The dataset includes details about gender, age group, customer satisfaction factors, shopping frequency, and shopping preferences. To ensure accuracy, I first cleaned the data by checking for any missing or NaN values. I also noticed that some individuals had multiple responses, which were clearly an error. To maintain data integrity, I removed any columns that contained these duplicate entries.

3.2. Comparison of Online vs. In-Store Shopping Preferences

First, I analyzed the percentage of customers who prefer shopping online versus in-person. To get a clearer picture of how these preferences vary across different age groups and genders, I also created a bar chart to visually represent the trends.

3.3. Shopping Frequency Trends

Next, I examined how often customers shop and analyzed their shopping frequency. To better understand how these habits differ across age groups and gender, I created a bar chart to visualize the patterns.

3.4. Demographic Influence on Shopping Preferences

Since we had data on individuals of different age groups and genders, I explored how these factors influenced their shopping preferences whether they preferred online or in-store shopping, how often they shopped, and their opinions on various aspects of the shopping experience. This included their thoughts on whether online shopping might eventually replace physical stores, as

well as their perceptions of website quality, product quality, security and privacy, convenience, and logistics.

To understand how different age groups and genders relate to shopping habits and other factors, I used Chi-Square tests. These tests helped determine whether there was a significant relationship between these variables. To assess the strength of these relationships, I also calculated the p-value. I conducted the Chi-Square test using the chi2_contingency function from the SciPy library. To make the findings more visually clear, I used bar charts to illustrate the patterns and trends in the data.

3.5. Factors Influencing Customer Satisfaction

I evaluated the relationship between key factors, like website quality, security and privacy, product quality, logistics, and convenience and shopping preferences using Chi-Square tests and p-values. This helped me understand how these factors influence purchasing preference. To show how different age groups and genders responded to these factors, I created bar charts for a clearer visual representation of the trends.

3.6. Data Visualization and Interpretation

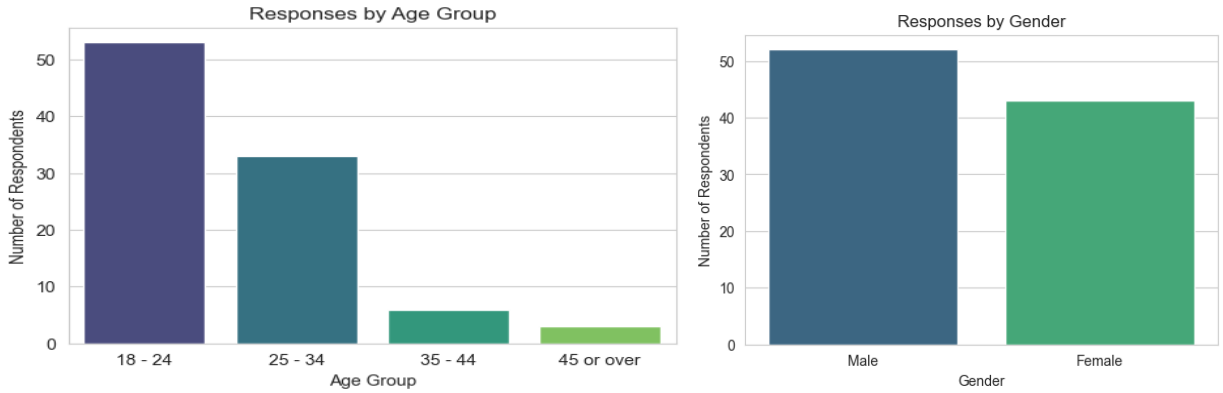
I used Matplotlib and Seaborn to create clear and engaging visualizations of the data. This approach helped me explore consumer shopping preferences, uncover meaningful trends, and draw valuable insights from the information.

4. Results and Discussion:

The analysis examined various factors influencing customer satisfaction in online shopping, based on survey responses. The key findings are as follows:

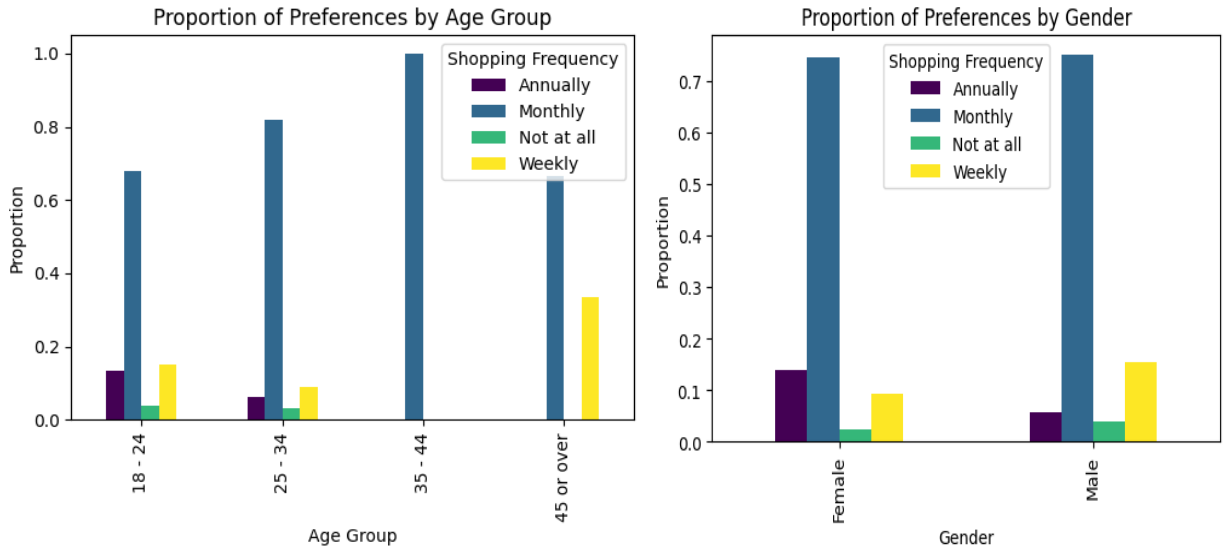
4.1. Demographic Insights:

The dataset reflects a diverse mix of participants across different age groups and genders. Among them, 48 are young adults aged 18-24, 32 fall within the 25-34 age range, 6 are between 35 and 44, and 3 are 45 or older. When looking at gender distribution, there are slightly more men (48) than women (41) in the respondent pool.



These demographic factors could shape the results, especially if there are notable differences in responses based on age or gender. The following details will take a closer look at these patterns and what they might mean.

I began by examining how often people of different **ages and genders go shopping**. The results show that most prefer to shop monthly, while a smaller group shop weekly. In both cases, the high p-value (well above 0.05) suggests no significant link between shopping frequency and either age or gender. However, the slightly lower chi-square value for age groups and shopping frequency suggests that age may be a slightly better predictor than gender in this context.

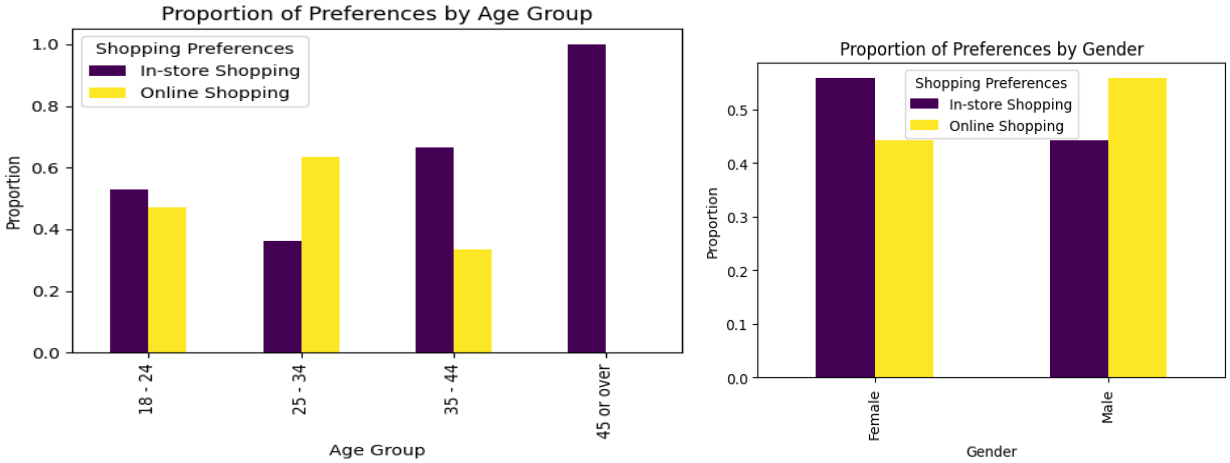


```
Observed Values:
 Age                        18 - 24  25 - 34  35 - 44  45 or over
How often do you shop online?
Annually                         7        2        0           0
Monthly                         36       27        6           2
Not at all                       2        1        0           0
Weekly                           8        3        0           1

Expected Values:
 Age                         18 - 24    25 - 34   35 - 44   45 or over
How often do you shop online?
Annually                   5.021053   3.126316  0.568421    0.284211
Monthly                   39.610526  24.663158  4.484211    2.242105
Not at all                 1.673684   1.042105  0.189474    0.094737
Weekly                     6.694737   4.168421  0.757895    0.378947

Chi-Square Value: 5.834673244984167
Reduced Chi-Square Value: 0.648297027220463
p-Value: 0.7563441194385945
Degrees of Freedom: 9
Conclusion: No significant relationship found.
```

```
Observed Values:
 Gender                     Female  Male
How often do you shop online?
Annually                        6     3
Monthly                        32    39
Not at all                      1     2
Weekly                          4     8

Expected Values:
 Gender                      Female       Male
How often do you shop online?
Annually                  4.073684   4.926316
Monthly                  32.136842  38.863158
Not at all                1.357895   1.642105
Weekly                    5.431579   6.568421

Chi-Square Value: 2.526854628066551
Reduced Chi-Square Value: 0.8422848760221836
p-Value: 0.4704573942115472
Degrees of Freedom: 3
Conclusion: No significant relationship found.
```

I then investigated consumer preferences for **online versus in-store shopping** based on age and gender. The results show that while most age groups tend to prefer in-store shopping, those aged 25-34 are more inclined toward online shopping. Among younger shoppers, the difference between in-store and online preferences is minimal, but slightly more still choose in-store shopping. However, older consumers, particularly those over 45, clearly favor shopping in physical stores. Regarding gender, men are generally more likely to prefer online shopping, while women tend to favor in-store shopping.

From a statistical perspective, the model appears to be underfitting, as the reduced chi-square value for age group and shopping preference is greater than one, indicating it does not fully explain the data. However, the p-value of 0.09, though slightly above the 0.05 threshold suggests, there could be a potential link between age and shopping preference. In contrast, while the chi-square results indicate a good model fit for gender and shopping preference, the high p-value suggests no significant relationship between these two factors.

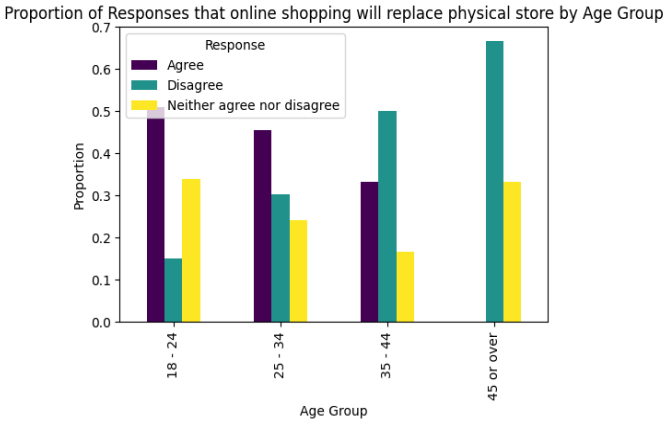


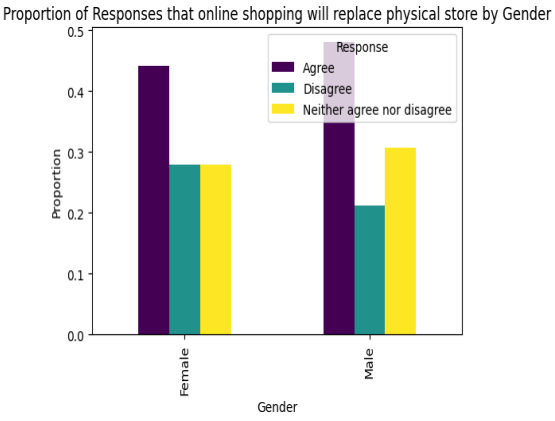
```
Observed Values:
 Age                18 - 24  25 - 34  35 - 44  45 or over
Preferences
In-store Shopping       28       12        4           3
Online Shopping         25       21        2           0

Expected Values:
 Age                  18 - 24    25 - 34   35 - 44  45 or over
Preferences
In-store Shopping  26.221053  16.326316  2.968421    1.484211
Online Shopping    26.778947  16.673684  3.031579    1.515789

Chi-Square Value: 6.2811931032525194
Reduced Chi-Square Value: 2.0937310344175066
p-Value: 0.09870284571284414
Degrees of Freedom: 3
Conclusion: No significant relationship found.
```

```
Observed Values:
 Gender              Female  Male
Preferences
In-store Shopping       24    23
Online Shopping         19    29

Expected Values:
 Gender                Female       Male
Preferences
In-store Shopping  21.273684  25.726316
Online Shopping    21.726316  26.273684

Chi-Square Value: 0.8424292822003578
Reduced Chi-Square Value: 0.8424292822003578
p-Value: 0.3587029179530338
Degrees of Freedom: 1
Conclusion: No significant relationship found.
```

Third, I investigated how consumers of different ages and genders feel about the possibility of **online stores replacing physical stores in the future**. The findings show that younger people, particularly those under 34, are more likely to agree with the idea that online shopping will eventually take over, while older individuals tend to disagree. When it comes to gender, both men and women generally agree with this statement.

The reduced chi-square value being close to one suggests that the model for age groups and their responses is slightly better from a statistical perspective. However, the p-value is higher than 0.05, indicating that there isn't a significant link between age and the belief that online shopping will eventually replace physical stores. Similarly, the large p-value for gender suggests no meaningful connection between gender and this opinion, with the chi-square results indicating that the model for gender and shopping preferences is overfitted.

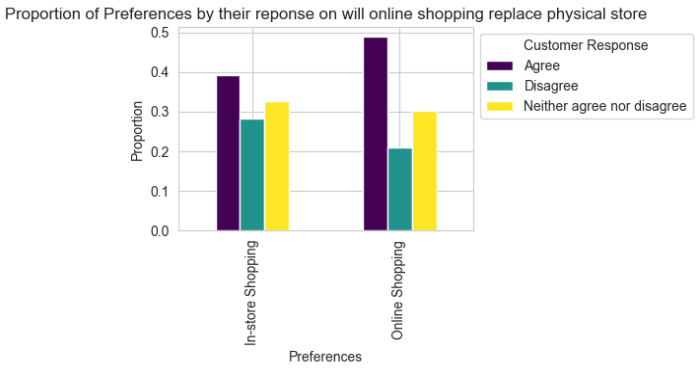




```
Observed Values:
 Age                                                  18 - 24  25 - 34  35 - 44  45 or over
"Online shopping will replace physical stores in the future"
Agree                                                     27       15        2           0
Disagree                                                   8       10        3           2
Neither agree nor disagree                                18        8        1           1

Expected Values:
 Age                                                    18 - 24    25 - 34   35 - 44  45 or over
"Online shopping will replace physical stores in the future"
Agree                                                24.547368  15.284211  2.778947    1.389474
Disagree                                             12.831579   7.989474  1.452632    0.726316
Neither agree nor disagree                           15.621053   9.726316  1.768421    0.884211

Chi-Square Values: 9.082970175100458
Reduced Chi-Square Value: 1.513828362516743
p-Value: 0.1689653767660813
Degrees of Freedom: 6
Conclusion: No significant relationship found.
```

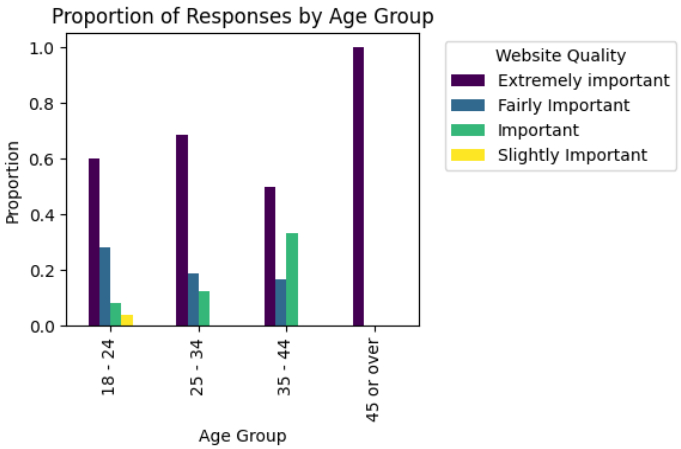
```
Observed Values:
 Gender                                               Female  Male
"Online shopping will replace physical stores in the future"
Agree                                                    19    25
Disagree                                                 12    11
Neither agree nor disagree                               12    16

Expected Values:
 Gender                                                 Female       Male
"Online shopping will replace physical stores in the future"
Agree                                                19.915789  24.084211
Disagree                                             10.410526  12.589474
Neither agree nor disagree                           12.673684  15.326316

Chi-Square Values: 0.5857138942957952
Reduced Chi-Square Value: 0.2928569471478976
p-Value: 0.7461288689258865
Degrees of Freedom: 2
Conclusion: No significant relationship found.
```

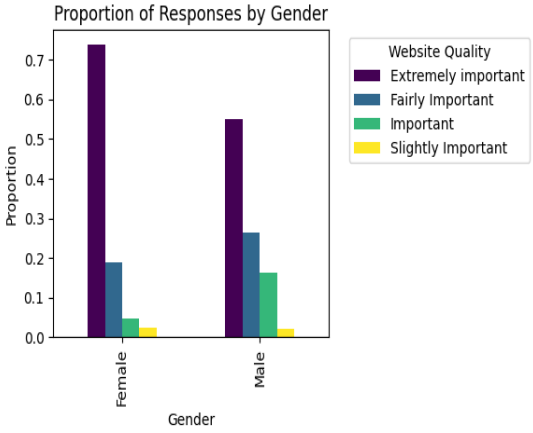
I also explored how people who **prefer online shopping** compare to those who prefer in-store shopping when it comes to the idea of **online shopping replacing physical stores**. The bar chart shows that most in-store shoppers also believe that online shopping will eventually take over, even though the chi-square and p-value don't reveal a significant connection between these factors.



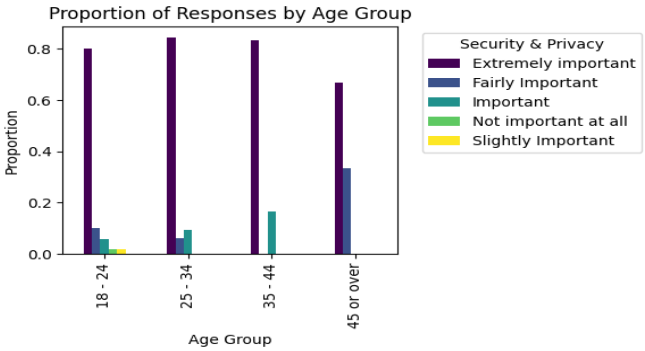
4.2. Key Factors Affecting Customer Satisfaction:

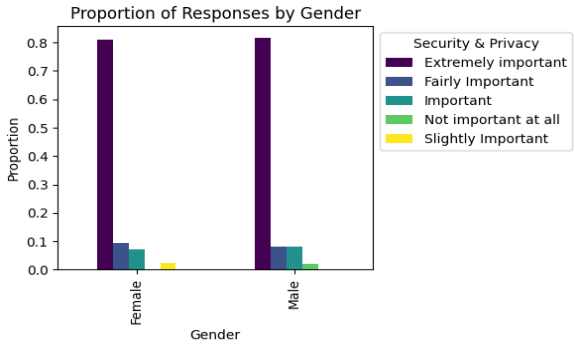
Website Quality: Many participants pointed out how important website design, ease of use, and navigation are to their shopping experience. I looked at their responses by age group and gender using a bar chart, and the results show that, no matter their age or gender, most people agree that a well-designed website is key.



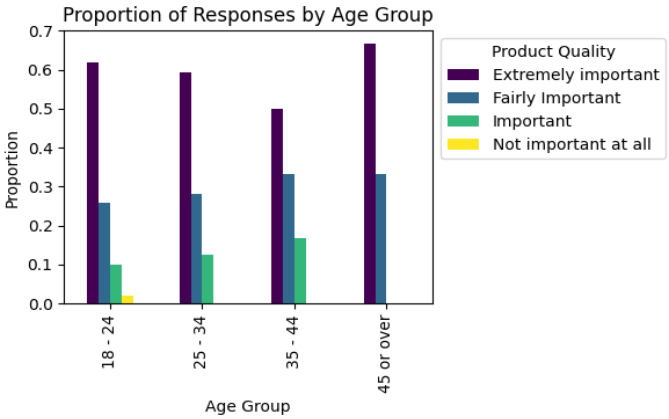


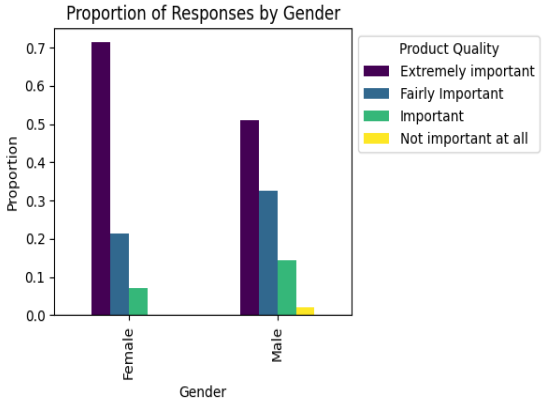
I also investigated how important **privacy and security** are for customer satisfaction, considering both gender and age. The results showed that, across the board, all customers, no matter their age or gender, view security and privacy as top priorities. While opinions on other factors were more mixed but over 80% of respondents agreed that security and privacy are the most crucial elements for customer satisfaction.



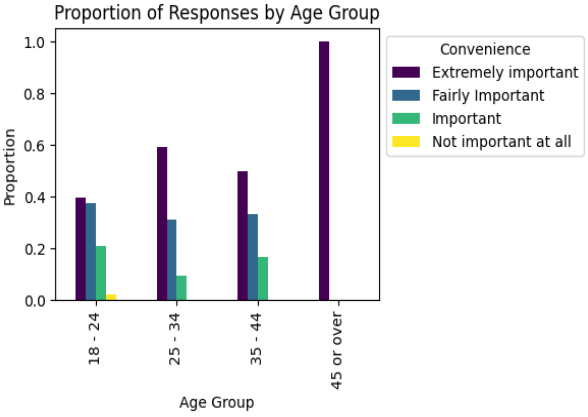


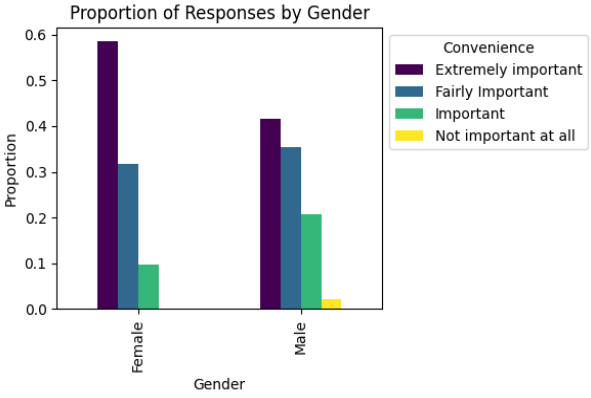
Across all age groups, consumers agree that **product quality** is key to customer satisfaction, which aligns with the overall findings. However, those over 45 tend to rate it as "extremely important" more often. When looking at gender, both men and women recognize the importance of product quality, but men's responses are more varied. In contrast, over 70% of women chose "extremely important," showing that female customers place a stronger emphasis on product quality.

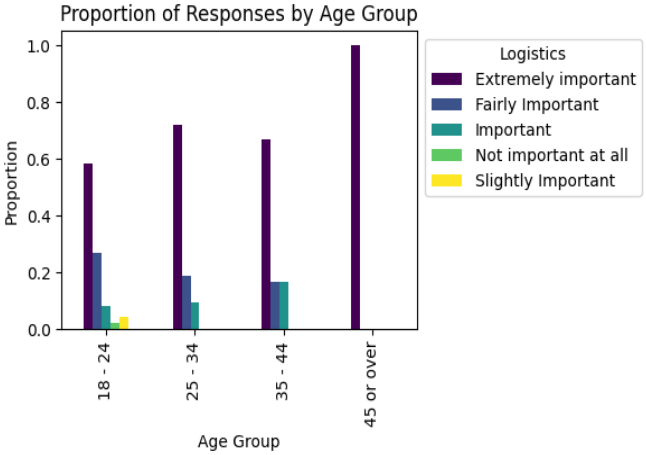


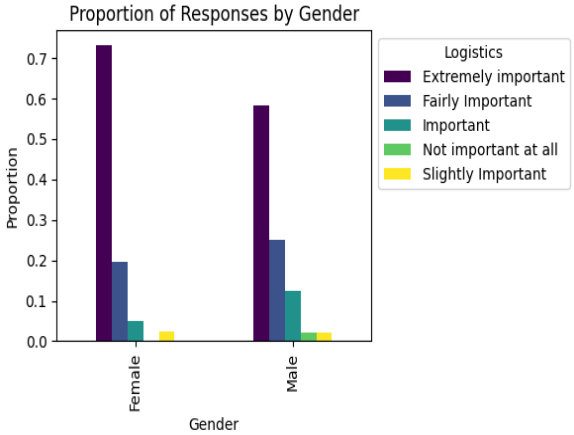


I also looked at **convenience and logistics** as key factors. The analysis by age and gender shows that all customers consider them highly important, with those over 45 placing even more emphasis on their significance. However, it's worth mentioning that most of our respondents are younger, with most being under the age of 34.









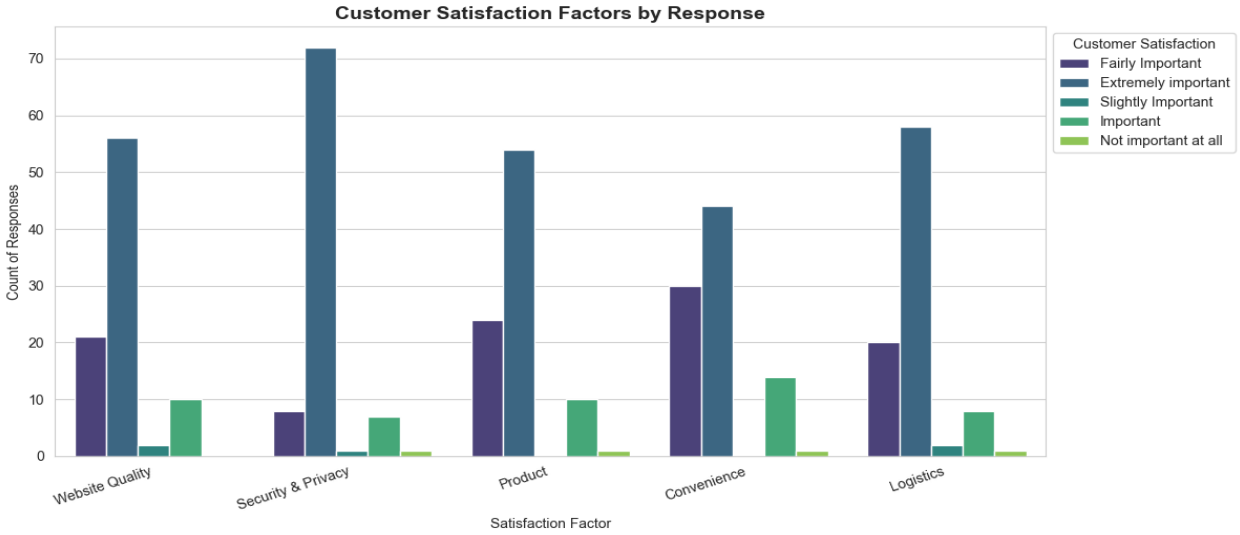
In addition to creating a cumulative graph showing the **number of responses** for each of **the five customer satisfaction factors**, I also ran chi-square and p-value tests on age and gender in relation to the responses. The results show that the model is overfitting the data, as the reduced chi-square value for the

relationship between age groups and these factors is much lower than 1. This, combined with a p-value greater than 0.05, suggests that there's no significant correlation between age and these factors. However, the model fits the data well when considering gender, as the reduced chi-square value is close to 1.

```
Reduced Chi-Square Value: 0.4566428297432278
p-Value: 0.9999358612210958
Degrees of Freedom: 63
Conclusion: No significant relationship found.
```

```
Reduced Chi-Square Value: 0.9217465644953856
p-Value: 0.5622711387198975
Degrees of Freedom: 21
Conclusion: No significant relationship found.
```

Additionally, the bar chart below shows how customers generally responded to these five factors. It highlights that most consumers consider privacy and security to be the most important aspects of online shopping.



4.3. Conclusion

In conclusion, the results show that while age and gender do have some influence on shopping habits, they do not seem to significantly affect key factors like how often people shop or whether they believe online stores will eventually replace physical ones. Across all age groups, both younger and older consumers place high importance on factors like website quality, convenience, and product quality. Security, privacy, and logistics also stand out as top priorities for everyone, with just slight differences between groups. To boost customer satisfaction and stay ahead in the competitive online retail market, e-commerce platforms should focus on enhancing their website experience, security measures, product quality, and logistics.

5. Critical Analysis

5.1. Statistical Limitations & Overfitting

The findings suggest that in some cases, the models used in the analysis, particularly those related to gender and shopping habits may be overfitting. This means that instead of identifying meaningful patterns, the models might be picking up on random variations in the data, making them less reliable for predicting real-world trends. Additionally, the fact that some factors have very low reduced chi-square values (well below 1) indicates that the model may not fully capture the diversity of responses, potentially missing important details and variations in consumer behavior.

5.2. Weak Statistical Significance

Many of the relationships analyzed had high p-values (>0.05), indicating no strong statistical connection between factors. While a p-value of 0.09 (for age and shopping choice) suggests a potential link, it still exceeds the conventional 0.05 threshold for statistical significance. This means the results should be viewed with caution, as the evidence is not strong enough to confirm a definite relationship.

5.3. Demographic Bias

Since most of the respondents are under 34, the dataset is skewed toward younger people. As a result, the views of older customers are underrepresented, which could affect how widely the findings can be applied. Additionally, there is a slight imbalance in gender, with more men (48) than women (41), which could introduce some gender-related biases into the results.

5.4. Consumer Behavior Beyond Demographics

Given the weak statistical relationships shown between age/gender and shopping preferences, it is possible that other variables, such as money, digital literacy, and personal habits, have a greater influence on purchasing behavior than demographics. To gain a better understanding, future studies should consider factors like income, education level, and familiarity with technology.

5.5. Business Implications

The strong focus on website quality, security, and logistics shows that these should be top priorities for e-commerce businesses, even though age and gender don’t have a major impact on buying behaviors Companies should focus on personalization strategies that cater to consumer preferences individually rather than relying solely on demographic-based marketing.

6. References

- Pandas
  McKinney, W. (2010) *Data Structures for Statistical Computing in Python*. Proceedings of the 9th Python in Science Conference, pp. 51–56.

- SciPy (chi-square test - scipy.stats.chi2_contingency)
  Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., et al. (2020) 'SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python', *Nature Methods*, 17(3), pp. 261–272.
  https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.chi2_contingency.html

- Matplotlib
  Hunter, J.D. (2007) 'Matplotlib: A 2D Graphics Environment', *Computing in Science & Engineering*, 9(3), pp. 90–95.

- Seaborn
  Waskom, M.L. (2021) 'Seaborn: Statistical Data Visualization', *Journal of Open Source Software*, 6(60), p. 3021.