Shoe Survey Snapshot: Uncovering Purchase Patterns

Team Members - Group 08:
Aryan Kumar, Neeharika Kamireddy, Rajashree Ramaprabu, Saumya Anand, Tanvi Sheth

Introduction

In the world of footwear, countless brands vie for our attention. But what exactly compels us to choose one shoe over another? While comfort, style, and functionality undoubtedly play a crucial role, this project explores whether brand popularity, independent of any specific features, influences our shoe purchasing decisions.

Research Question

This project aims to understand the power of brand logos and their associated popularity. We'll investigate whether the mere presence of a well-known brand logo sways the decision-making process of our participants, even when all the shoes' functional characteristics remain identical. By disentangling the influence of brand popularity from the inherent qualities of the shoes themselves, we can gain valuable insights into the psychology behind consumer behavior and brand appeal in the footwear industry.

If our hypothesis is confirmed, it suggests that brand popularity, as signified by logos, holds considerable sway over consumer choices. Such insights can inform marketing strategies and product positioning within the industry, opening avenues for further research and strategic development. Conversely, if the hypothesis is not supported, this prompts a reflection on the nuanced factors influencing consumer decisions, encouraging a deeper exploration of the dynamics at play.

Null Hypothesis:

H0: The presence of a brand logo does not lead to a statistically significant difference in users' purchase propensity.

Alternative Hypothesis (H1): The presence of a brand logo on a shoe impacts user preferences and purchase intentions, irrespective of the shoe's inherent qualities.

Methodology

Survey Design:

For the survey, we selected a set of five shoes[1] that would be shown to each of the participants and the respondents were asked to answer a series of questions for each shoe. The participants were prompted to evaluate the shoe by assigning ratings solely based on its visual attributes. The participants were divided into two groups - control group, which displayed a plain shoe and the

[1] Image References: https://drive.google.com/drive/folders/1I2q84YpLJQZBMClj_w5l3XTi62dq1vKB?usp=drive_link

treatment group, which included the brand logo in the image. The questions we asked each participant for each shoe were:

1. Please rate the following shoe based on your preference.
2. How comfortable does the shoe appear?
3. Out of the features mentioned, which is the most important to you while buying a pair of shoes?
4. The shoe is priced in the range of $80-$100. How likely are you to consider purchasing it?

It is imperative to note that the above four questions remained constant across all the five shoes in the survey as well as across the control and treatment groups. The questions followed a five-point LIKERT scale with the only exception being question three which gave the respondents choices between features such as style versatility, material, durability, breathability, traction and grip. All of the questions were categorical, which gave us the opportunity to check the level of each variable we were checking.

By comparing the responses of the treatment group, exposed to the brand logo, with those of the control group, we aimed to identify any statistically significant differences. Utilizing appropriate statistical methods, we assessed the strength and direction of these differences, providing a robust foundation for drawing meaningful conclusions.

Participants

Our study involved 120 participants selected based on their age group. The participants are equally divided into two groups. Our participants consisted of students from different universities in the United States falling under the age group of 18 to 30 years of age. This diverse sample aims to capture a wide representation of potential customers of the footwear industry. By considering factors such as gender, and shopping habits, we strive to enhance the external accuracy of our findings. The inclusion of a significant sample size would support the reliability and generalizability of our results.

Randomization

In our experiment, we ensured to have an equal treatment by including the randomization unit. The experiment was randomized at the participant level. We randomly assigned people to either the group that saw shoes with a popular brand logo or the group that saw shoes without any logos. This randomization helps to avoid any unfair advantages and to provide a balance between the two groups. The idea behind the randomization is to make sure that any differences we observed in the treatment/control group's reactions could be linked specifically to the difference in the presence or absence of the brand logo and to ensure no other factors interfere with our results.

Data Collection:

To collect data for our experiment, we circulated the survey (displayed in Appendix A) among college students across all universities in the United States. The participants all belonged to the age group of 18 to 30 years of age.

To understand the influence of brand logos on consumer decisions, participants were divided into two groups: the treatment group and the control group. Both groups were presented with the same image of a shoe and in the same order, but with a key difference:

- Treatment Group: The image of the shoe featured a prominent brand logo.
- Control Group: The shoe presented to them was shown from a different angle, omitting any brand logos altogether.

Following this exposure, both groups were asked a series of four questions designed to gauge their perception and potential purchase interest in the shoe. By carefully controlling the only variable between the two groups - the presence of the brand logo - we aimed to isolate and analyze its potential influence on their decision-making process. This research approach helps us uncover the hidden power that brands may hold in shaping consumer preferences, even when other shoe features remain constant.

Data Analysis

Data Overview:

Our survey had 120 respondents, equally split between the control and treatment groups.

Columns	**Description**
Gender	Male or Female
Shoe_number	Shoe Number included in the survey - 1 to 5
Preference_Rating	Preference rating on a scale of 1 to 5 with 5 being the most preferred
Comfort	Perceived comfort rating on a scale of 1 to 5 with 5 being the most preferred
Important_Features	Style Versatility, Material, Durability, Breathability, Traction and grip
Likelihood_of_purchase	Purchase Propensity on a scale of 1 to 5 with 5 being the most preferred
Treatment	Control or Treatment group

Table 1: Data Overview

Data Exploration:

As part of our initial data analysis, we created five different graphs to compare and contrast the purchase patterns and key factors associated with that.

1. Likelihood of Purchase for Treatment and Control Groups (Fig. 2.1): We started with comparing the likelihood of the purchase between the two groups. The graph represents that the control group is more likely to purchase the shoe 1 with a count above 100. The same pattern appears for shoe 2. However for shoes 3, and 4, we can notice a steady growth in purchase opportunities from the treatment group. At shoe 5, we can notice the likelihood of both groups has dropped.
2. Distribution of Shoe Preference Ratings (Fig. 2.2): We have next made a graph to showcase the rating preference of each shoe based on the count of shoes. For shoe 1, the count is slightly higher than 140. For shoe 2, the count is around 110. Shoe 3 had a preference rating of about a count of 130. Shoe 4 has the highest range based on preference rating with a count range above 140. Shoe 5 has the lowest count with preference rating with a range of 60.
3. Comfort Level Preferences for Treatment and Control Groups (Fig. 2.3): In our next graph, we visualized how frequent each shoe is chosen based on the comfort level factor. We visualized this pattern separately for treatment and control groups. Shoes 3 and 4, are the most often chosen ones based on the comfort level by the students from the treatment group. The pattern of frequency of how a shoe is chosen based on the comfort it provides is similar across both groups, however the count of it varies drastically where the treatment group has a higher number and the maximum frequency number reached by the control group is slightly above 80.
4. Shoe Preferences based on Features (Fig. 2.4): The next graph visually covers shoe preferences based on features. For shoes 1 and 2 are highly preferred for their style versatility and secondly for their durability. Shoe 4 has a preference for style versatility and durability. Shoe 5 has an equal spread out preference for style versatility, durability and breathability. Overall, traction and grip and material are the least looked out features when purchasing for a group.

Estimated and Conditional Average Treatment Effect

The estimated average treatment effect for the entire population is 0.23. We can infer that the treatment has a positive effect on the purchase propensity. From the analysis, we discovered that the average treatment effect is notably different between genders: 0.38 for men and 0.06 for women. This suggests that while both genders exhibit a positive response to the treatment, men demonstrate a higher propensity than women to purchase shoes influenced by the brand.

Regression

Regression analysis is a valuable tool for understanding the connections between different factors and the likelihood of making a purchase. In the context of assessing the probability of purchase, we used regression to explore how variables like gender, preference rating, comfort rating, and important features related to variations in purchasing behavior.

Essentially, regression helped us uncover the strength and direction of these relationships. For example, we investigated whether gender significantly influenced the likelihood of a purchase, understood how shifts in preference or comfort ratings aligned with buying behavior, and pinpointed specific features that played a substantial role in shaping the decision to make a purchase. By employing regression analysis, we gained insights into the intricate dynamics among various independent variables and the likelihood of purchase. This information was crucial for making informed decisions and predictions based on the observed relationships within the data.

In our pursuit of understanding the factors influencing the likelihood of purchase, we conducted multiple regression analyses. We started by checking the most basic variable - treatment. We know from the calculation above that the expected ATE is positive, but we thought it better to dig deeper into whether it is statistically significant or not.

	Dependent variable: Likelihood of Purchase
	OLS w/o controls
	(1)
Constant	2.340^{***}
	(0.076)
Treatment	0.230^{**}
	(0.107)
Observations	600
R^2	0.008
Note:	*p<0.1; **p<0.05; ***p<0.01

Fig. 1: Regression analysis with Treatment

Examining the outcomes of the regression analysis, we ascertained that the treatment exhibits a statistically significant positive impact. With this confirmation of the treatment effect, our focus shifted to investigating how the various variables influence our target.

Initially, the analysis included important features as one of the independent variables. However, we observed a notably high p-value associated with the important features, indicating a lack of statistical significance. As a result, we decided to omit this variable from the model and re-evaluate the remaining variables.

Upon revisiting the analysis without the important features, our attention turned to gender, comfort, and preference as potential contributors. Surprisingly, the statistical tests revealed that gender did not emerge as a significant predictor. In contrast, both comfort and preference exhibited statistically significant relationships with the likelihood of purchase. These findings suggest that, in our model, the impact of gender on purchase likelihood was not discernible, while comfort and

preference played discernible and statistically significant roles in influencing purchasing behavior. These insights offer valuable guidance for refining our understanding of the key determinants affecting the likelihood of purchase in the context of our study.

Fixed Effect:

In our regression analysis, we incorporated the shoe number as a fixed effect. This approach accounts for potential variability associated with each specific shoe within the dataset, allowing us to control for individual shoe characteristics that may impact the target variable. By treating shoe number as a fixed effect, we enhance the precision of our analysis, ensuring that the observed effects are not solely attributed to inherent differences between individual shoes. This nuanced consideration enables a more robust examination of the relationships between the independent variables, such as treatment and other relevant features, and our target variable, which is the likelihood of purchase, further refining the accuracy and reliability of our findings.

Results

T-test:

The t-score from our experiment was 2.1518. This score is a measure of the difference between your two groups, relative to the variation observed in our sample data. A higher t-score indicates a greater difference between the groups. In this case, a t-score of 2.1518 suggested that there was a noticeable difference in the 'Likelihood of Purchase' between the treatment and control groups.

P-value:

The P-value was found to be 0.031816 (3.1816e-02, to be precise). This value helped us understand the significance of the results. In simpler terms, it indicated how likely it was to observe the data we had if there had been actually no real difference between the groups (as stated by the null hypothesis). Since our P-value was lower than the minimum threshold of 0.05, it suggested that the difference we observed was unlikely to have occurred merely by chance.

```
t-score (t): 2.1518
P-value (p): 3.1816e-02
Should we reject H0? Yes
```

Based on these results, and particularly the P-value, we concluded to reject the null hypothesis. This meant that we had sufficient statistical evidence to believe that the presence of a treatment led to a significant difference in the 'Likelihood of Purchase'.

Statistical Power:

In our analysis, we calculated the statistical power of the experiment to be 0.57. This measure indicated the experiment's ability to correctly reject the null hypothesis when it was indeed false. A power value of 0.57, which is slightly low, suggested that our experiment had a moderate

likelihood of detecting a true effect, in case one existed. In practical terms, this meant that while our experiment was somewhat capable of uncovering significant differences, there was still a notable chance of not detecting such differences if they were present. This level of power highlighted the need for cautious interpretation of the results, as the risk of failing to detect a true effect was higher than ideal.

Challenges and Limitations

One limitation we should acknowledge in our analysis is that our experiment was carried out exclusively with college students. Considering the heightened awareness among individuals in this age group regarding different types of shoes, it's possible that participants in the control group may have been able to identify the brand of the shoe without explicitly focusing on the logo. This unintended recognition might have influenced their responses, adding a layer of complexity to the interpretation of our results.

Another limitation that we experienced was the lack of statistical power caused due to insufficient number of samples. This limitation was caused due to the lack of time and resources. If done on a larger scale, we have the possibility to scale this experiment to all university students and get more insights into other reasons that would explain their purchase intentions.

Conclusion

Based on the findings from our concise yet insightful experiment, we were able to confidently reject the null hypothesis. This meant that the presence of a brand logo on a shoe image significantly influenced users' preferences and intentions to purchase, overshadowing the intrinsic qualities of the shoe itself. This outcome underscores the powerful impact that brand recognition and imagery can have on consumer behavior.

Further, our experiment revealed that not only does the brand logo play a crucial role, but also the visual appearance and perceived comfort of the shoe, as gauged from the image, are key determinants in forming a user's impression. These factors collectively contribute to shaping the potential buyer's perception, significantly swaying their decision-making process.

This demonstrates the complex interplay of brand image, visual aesthetics, and perceived functionality in influencing consumer choices. In the realm of marketing and product design, these insights are invaluable, highlighting the importance of not just the actual product quality, but also the power of branding and presentation in appealing to customers' preferences and buying decisions.

Appendix

Appendix A: Survey

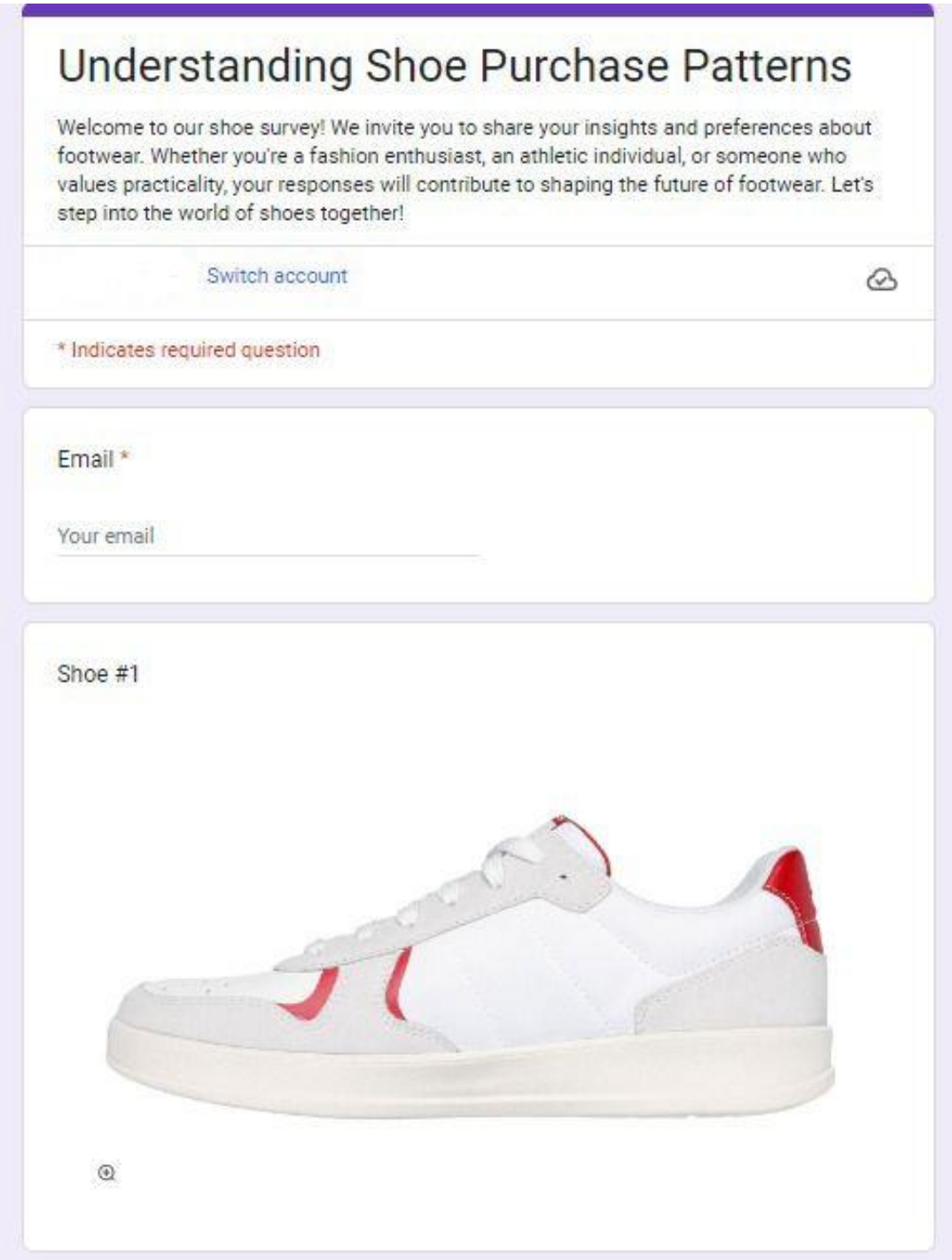

Image 1: Shoe Image for Control Group

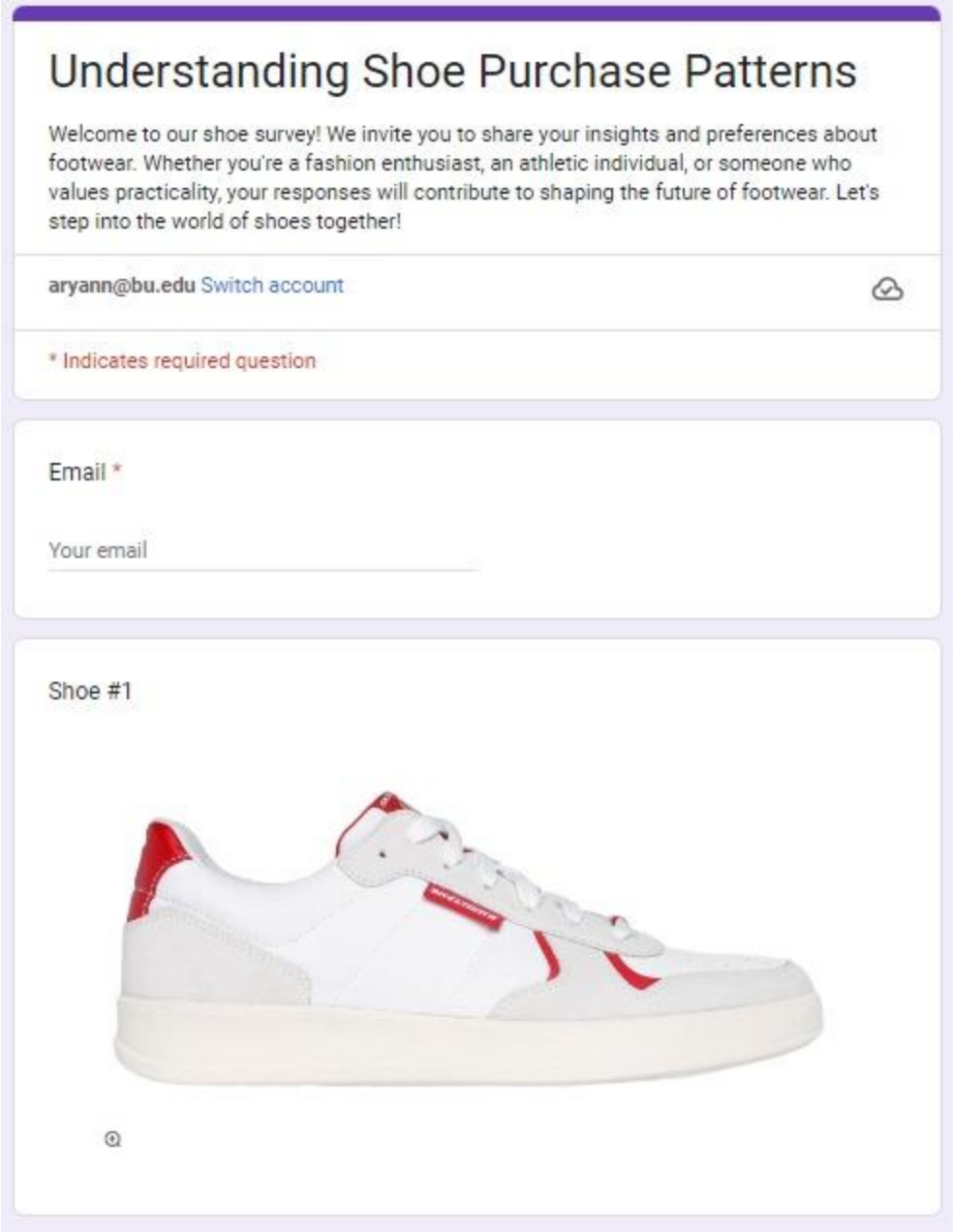

Image 2: Shoe Image for Treatment Group

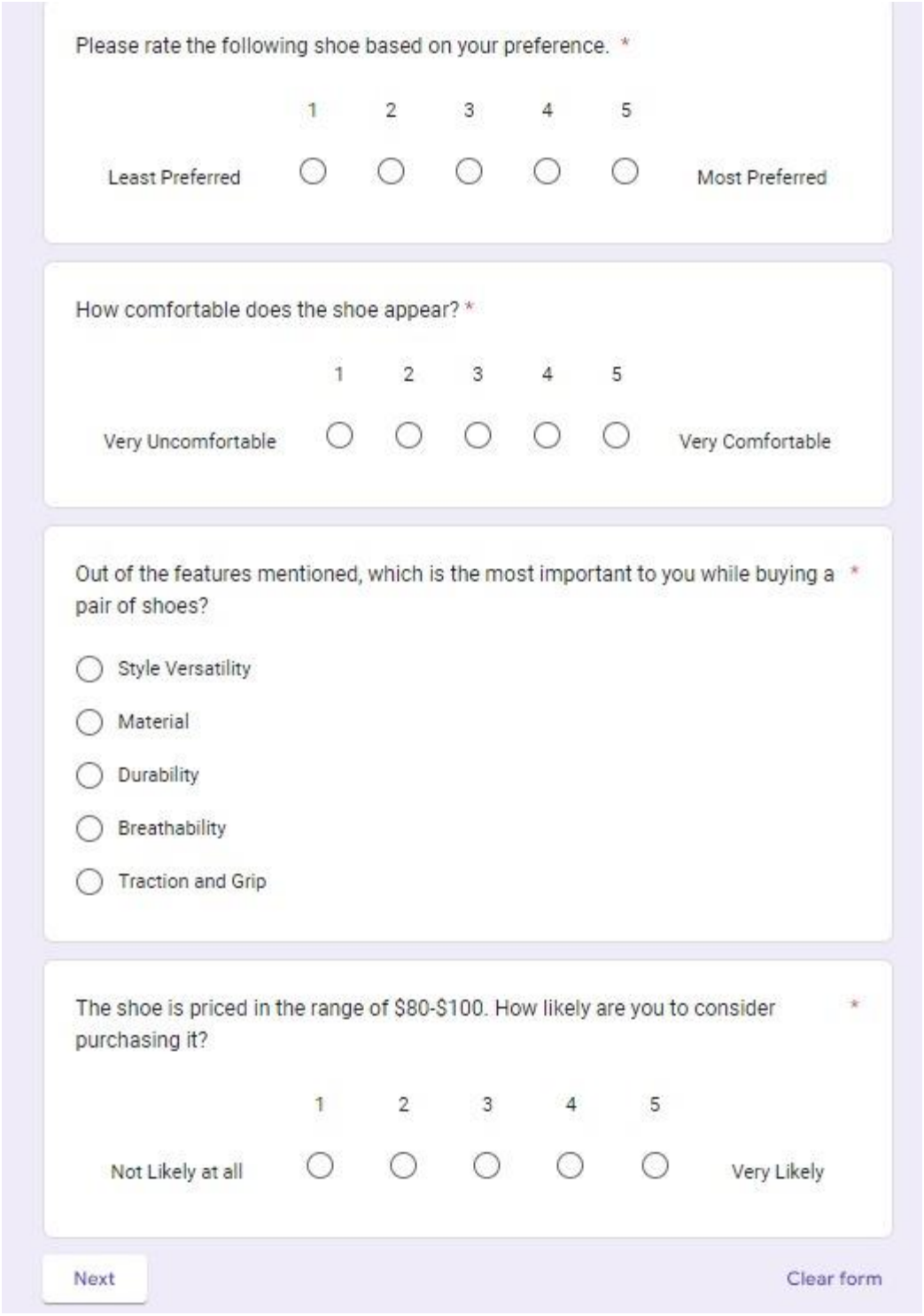

Image 3: Survey Questions

Appendix B: Data Exploration

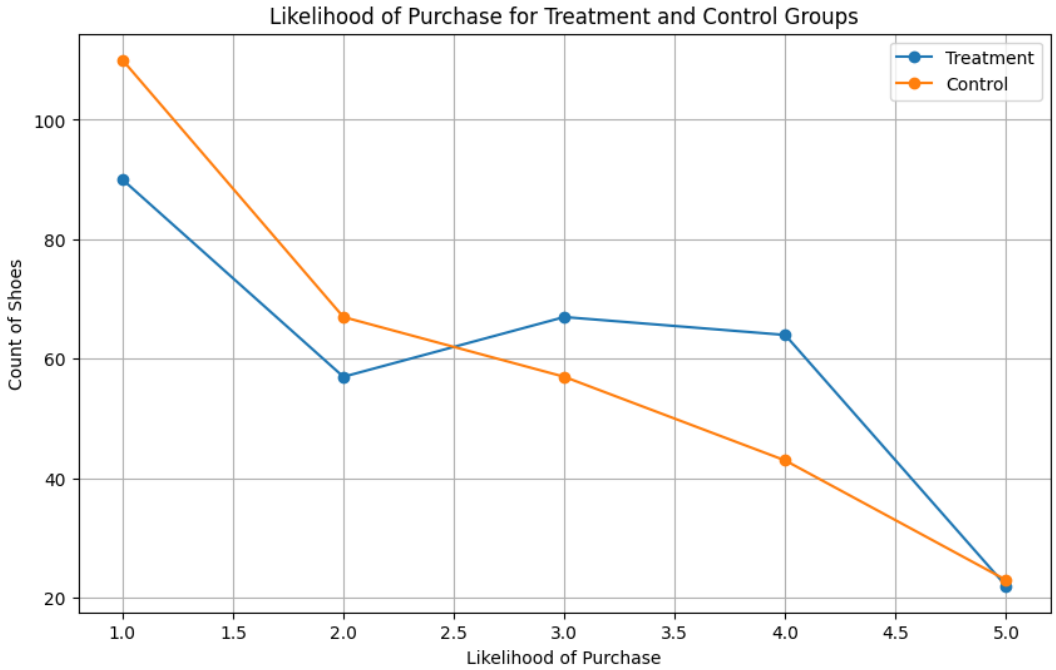

Fig. 2.1: Likelihood of Purchase

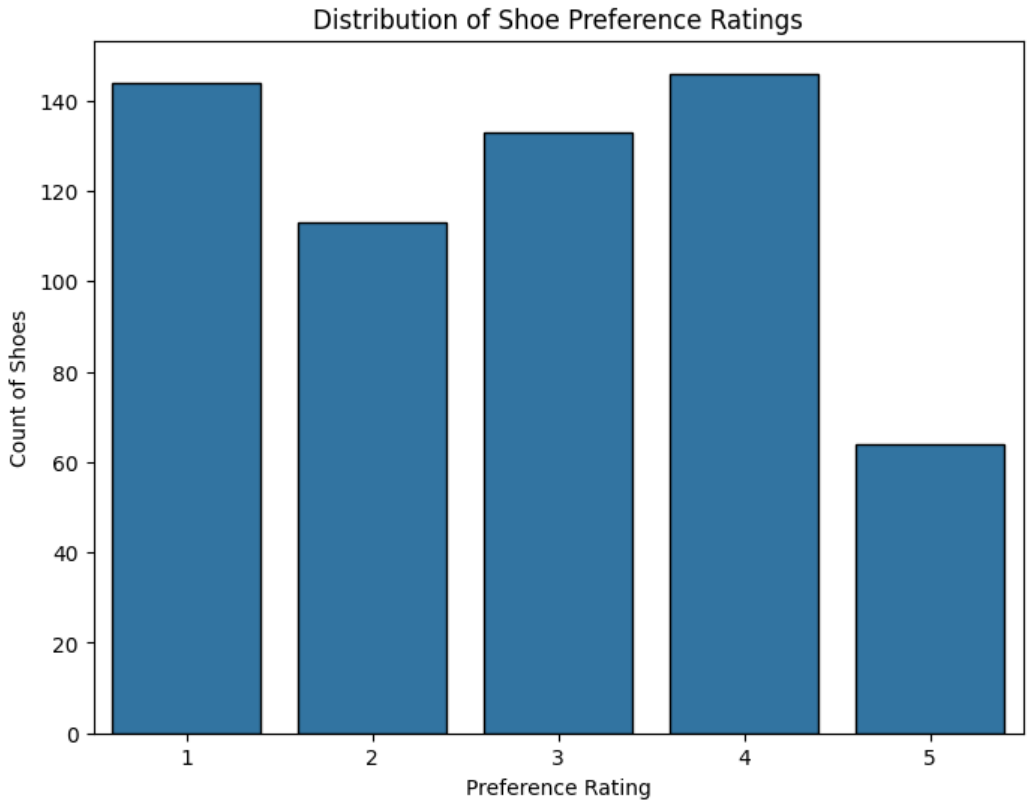

Fig. 2.2: Distribution of Shoe Preference Ratings

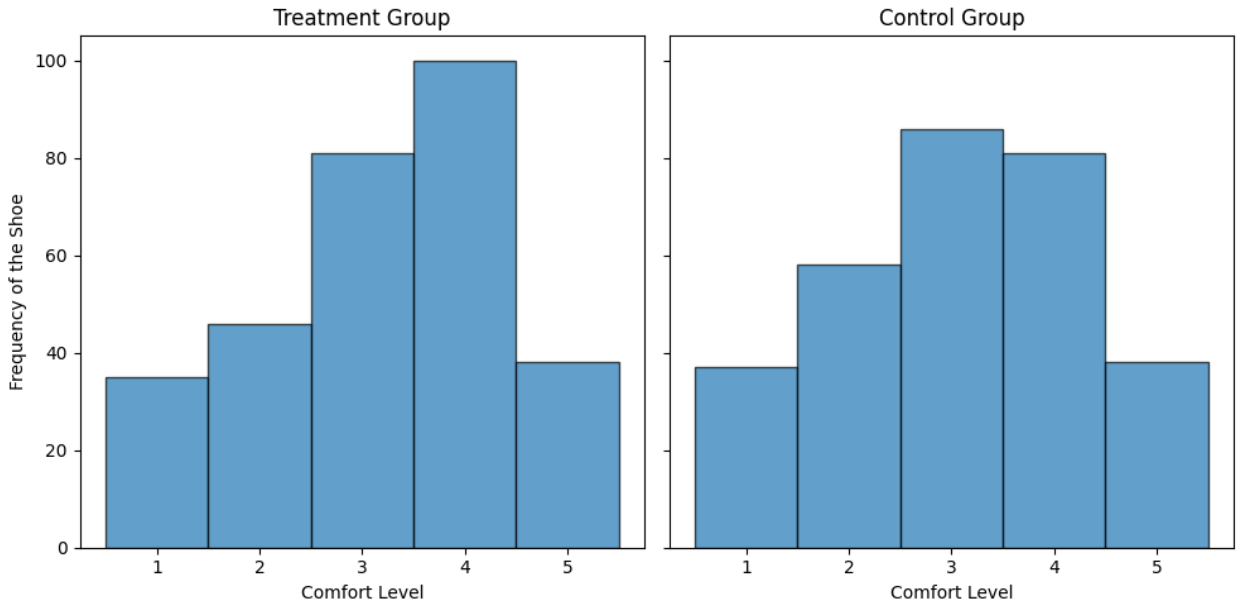

Fig. 2.3: Comfort Level Preferences for Treatment and Control Group

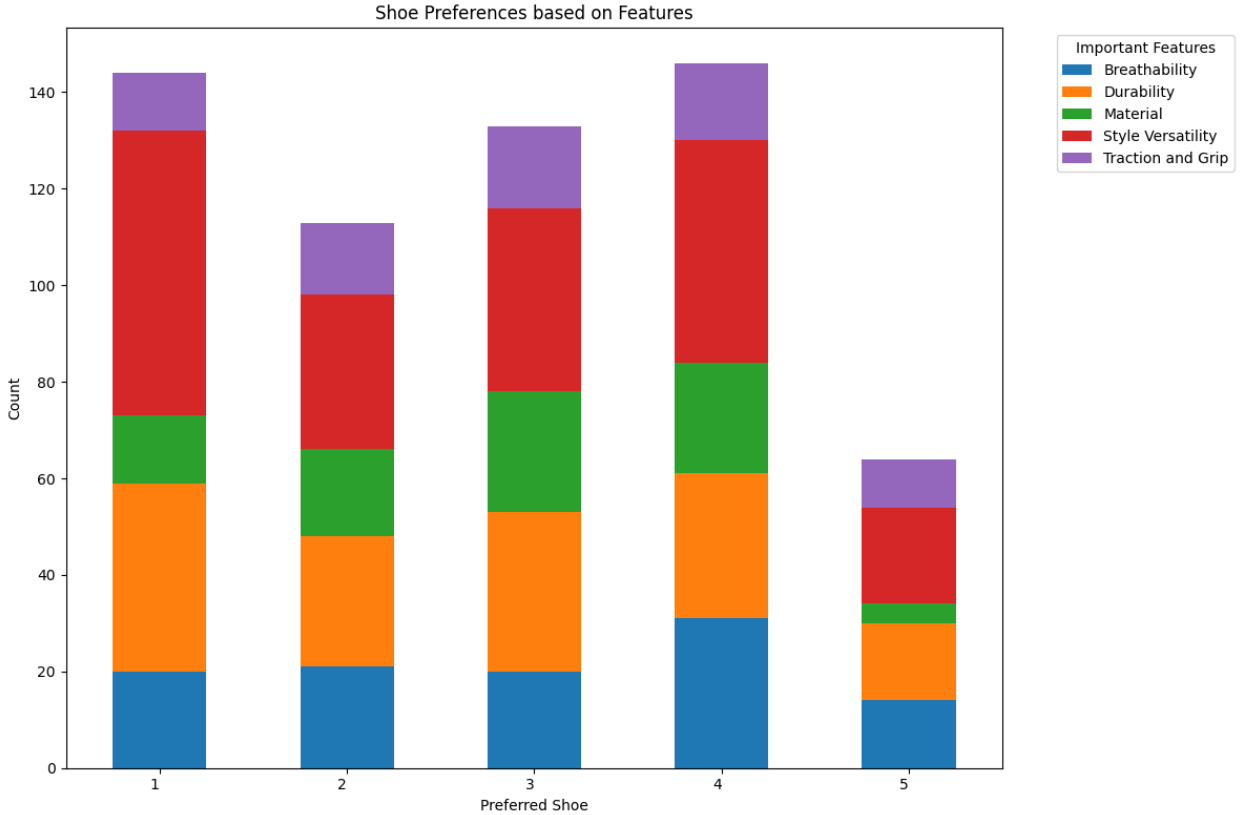

Fig. 2.4: Shoe Preferences based on Features