A/B Testing Final Project Report
Apple Music Subscription

Niharika Patil(niharikp), Apoorva Shetty(ashetty3), Hiba Hassan (hibah)
Carnegie Mellon University

Our Question

"Does the exposure of an individual to a coupon code for a free 3-month trial, along with information on the features of a music streaming platform within an advertisement, influence them to test, switch to, or develop a preference for the advertised platform over other music listening applications?"

Our main focus is to understand the behavior of Apple device users who do not primarily use Apple Music for streaming music. The information obtained from this study could shape future marketing strategies for Apple Music, aiming to establish it as a leading music streaming service. Additionally, this experiment will shed light on whether users' loyalty to their current platform, based on previous usage, affects their willingness to experiment with a new streaming service.

Experiment Design - A/B Test

Hypothesis

*Null***:** Exposure to the advertisement message detailing the 3-month free usage will not affect an individual's intention to:
1. Try the new streaming service
2. Switch to the new streaming service
3. Develop a preference (change in preference) for the streaming services

*Alternate***:** Exposure to the advertisement message detailing the 3-month free usage will affect a user's intention to do at least one of the above.

Data Collection

We used the Qualtrics Survey tool to conduct this experiment, following an A/B Test framework. We collected initial demographic information of the survey participants, along with their preferred primary music streaming service preference. We also collected a rank-ordered list of all the music services they use. Additionally, we obtained details on the features and attributes that the participants value in their chosen music streaming service.

Treatment

In our study, the intervention presented one group of survey respondents with a comprehensive message about the music streaming service, accompanied by a coupon code for a free three-month trial. Conversely, the control group received a concise message about the streaming service but did not receive a coupon code for complimentary access.

Randomization

Our primary interest was to understand the behavior of participants towards the marketing message for Apple Music. Hence, we subjected those participants who do not use Apple Music as their primary streaming service, to Apple Music advertisements. Since the sample size for this survey was quite small, we exposed the primary Apple Music users to advertisement messages for Spotify. The treatment and control messages were randomized among survey takers within these sub-groups.

Overall Outcome and Evaluation Criterion

To evaluate the difference in behavior among treatment and control groups, we use a success metric "**Positive Intent**". We define the value of this metric as "1" for individuals who display an intent to either try the advertised streaming service, switch to the advertised streaming service, or if the advertised streaming service moves up in their rank-ordered list of preferences of music services. If an individual does not change intent, we assign the metric a value of "0".

To estimate the difference across the treatment and control groups, we first average this value across both groups. We then perform t-tests using these averages and 90% and 95% confidence levels (90% significance level owing to the small sample size of the experiment).

We also want to study the difference in the averages across sub-strata like changes in behavior across Apple Product users and Non-Apple Product Users, music preferences, willingness to pay for music, consumer loyalty, and finally the platform they're already using. We supplement the overall analysis by running tests on the differences in means of each of the three metrics mentioned above for a more detailed analysis and evaluation of effects within groups.

Data Analysis and Methodology

We carried out some exploratory data analysis to understand the survey and responses a bit better.

- Overall, we received 48 survey responses, out of which 13 (27%) responders used Apple Music as their primary streaming service and 35 (73%) did not.
- Most of the respondents 20-30 years old and 81% (39) of them use iPhones
- The breakdown of Iphone Owners 80% (39) vs Android Owners 15% (7)

- *Age and Phone Type:* A major chunk of respondents owned an iPhone and were between the ages of 20-25 representing a relatively young set of respondents.
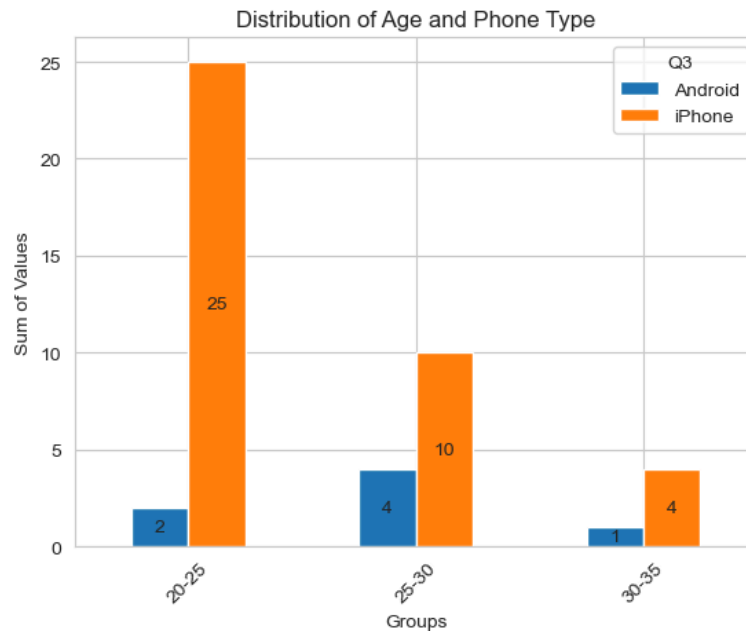

- *Age and App Preference:* Spotify was the more popular choice across all age groups.
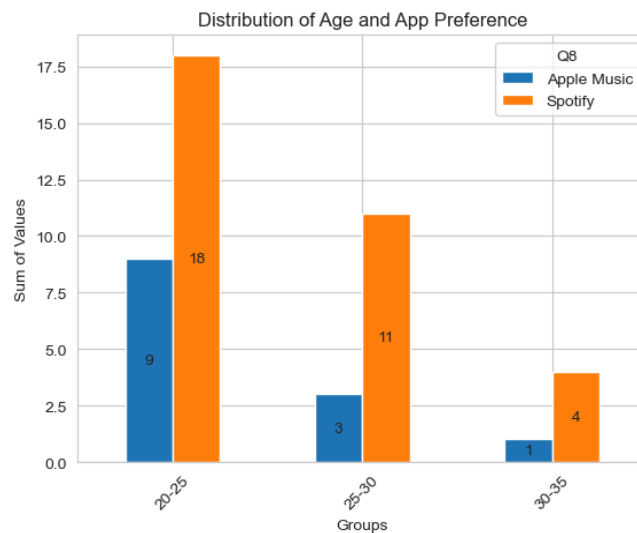
- *Distribution of willingness to spend vs phone type:* We observe that the majority of respondents are willing to spend between $5-$10 and $10-$25 for a streaming service.
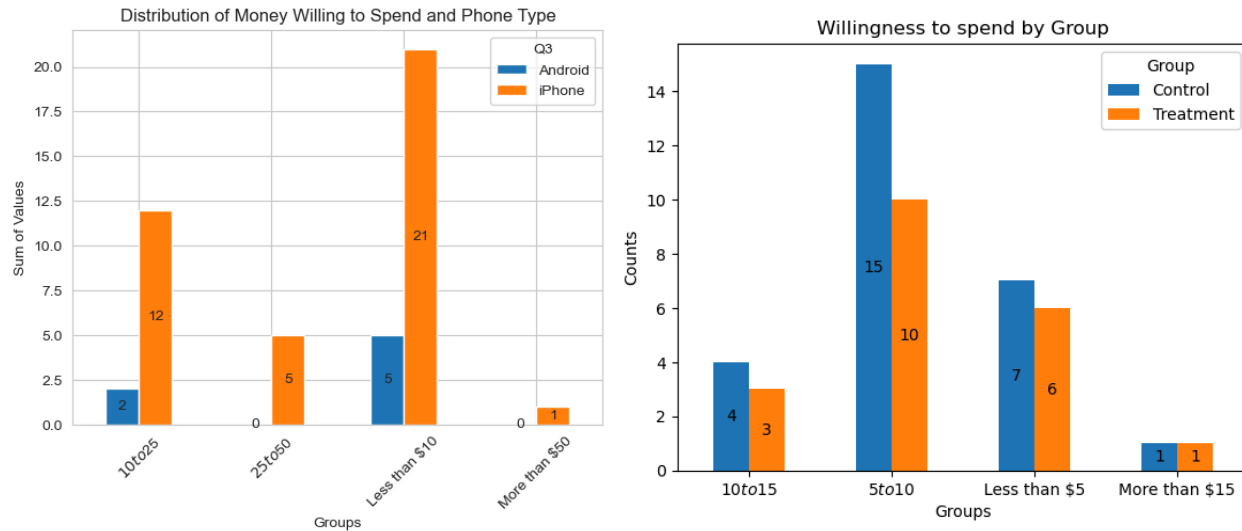
- *The Apple Ecosystem:* We wanted to analyze how the idea of 'apple ecosystem' could potentially trickle down into preferring one streaming service over another
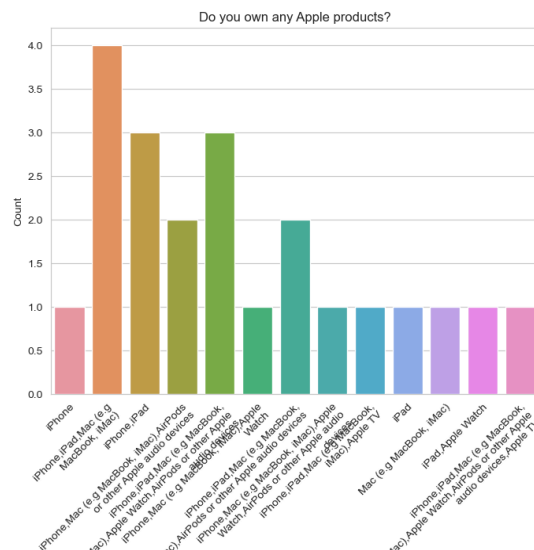
Overall our visualization analysis helped us understand how demographic differences could potentially affect how the treatment processes through. Given that there is a higher number of Iphone users, it can be natural to expect that it is easier for majority respondents to be more responsive to what our experiment entails. Furthermore, the fact that the majority does prefer Spotify over Apple Music gives us more concrete reasoning to push this experiment through as a

potential sway of consumers towards Apple Music. Finally, we are able to identify that many users are not willing to pay beyond $10 which is interesting as that could be a reason for their lack of interest in Apple Music subscription itself, pre-treatment.

Results and their Interpretation

Overall Results

Metric	Observed Difference	T-statistic	Significance
Cumulative Positive Intent	-24.29%	-1.97	Significant (90%, 95%)
Positive Intent to Try the Advertised Streaming Service	-27%	-2.11	Significant (90%, 95%)
Positive Intent to Switch to the Advertised Streaming Service	-23.68%	-1.65	Insignificant (90%, 95%)
Positive Intent Measured by Change in Preference	4.45%	0.5424	Insignificant (90%, 95%)

Interpretation
- The campaign had a statistically significant negative impact on the intent to try the advertised music streaming service, which is a critical concern. It suggests that something about the campaign or the service as presented may have actively discouraged potential users.
- The lack of significant effects on the intent to switch and the change in preference rankings suggests that the campaign did not effectively persuade existing users of other services to consider switching or altering their preferences in favor of the advertised service.
- The overall effectiveness of the advertisement in discouraging positive perceptions or actions towards the music streaming service is questioned, particularly given the significant negative impact on the willingness to try the service. This indicates a need to reassess the campaign's messaging, targeting, or perceived value of the service to better align with potential users' expectations and preferences.

Stratified Results

We worked on identifying stratas within the treatment and control groups to align our analysis in a more coherent way. Our approach is as follows:

Hours of streaming music and podcasts

Key Takeaway: We observe that people in the control group spent more time streaming music overall as compared to those in the treatment group.

For more effective analysis, we segmented the people into two sub-groups, people where the Low to Moderate group contained people who listened to music for up to 10 hours a week, and the rest were categorized as 'High'. When we examined each sub-group the only statistically significant difference was in the intent of people who listen to a low-moderate amount of music per week in the control group wanting to switch to the recommended streaming service. We observed a 35.7 percentage point difference in means with a p-value of 0.048 on a 5% confidence level. The lower mean suggests that the treatment is not effective in influencing the decisions of this subgroup.

Supporting Data Visualization and Table:
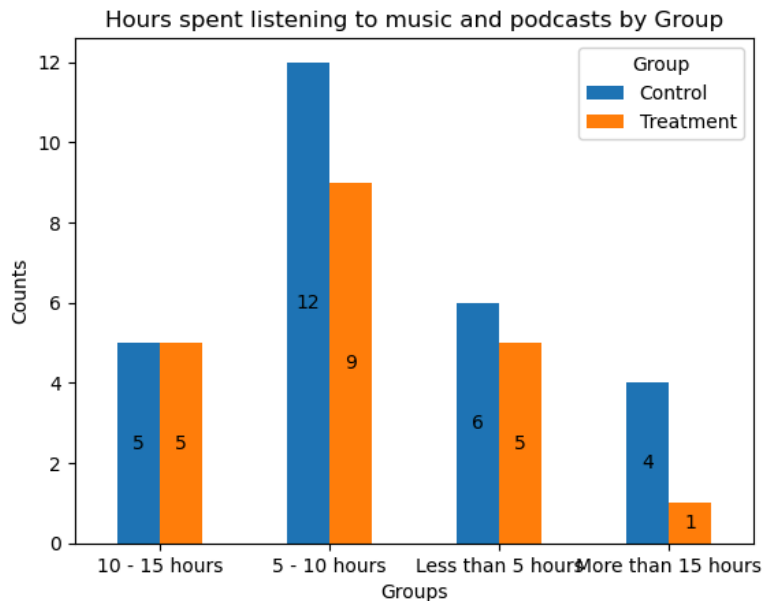
Metric	Value
Difference in Averages	-0.357
T-Statistic	-2.05
P-Value	0.048

Understanding Consumer Loyalty

Key Takeaway: We observed that there was a disproportionately high representation of loyal customers in the treatment group as compared to the control group.

This would indicate that the treatment group would be less likely to change their mind about their preferred streaming service as compared to the control group hence putting a downward bias on the overall results. When we looked deeper into the three metrics of success, there wasn't enough data available to justify a significant difference in means so we did not pursue further analysis on this matter. The t-test to evaluate the difference between these groups was insignificant at -0.75 on a 5% confidence level.

Supporting Data Visualization
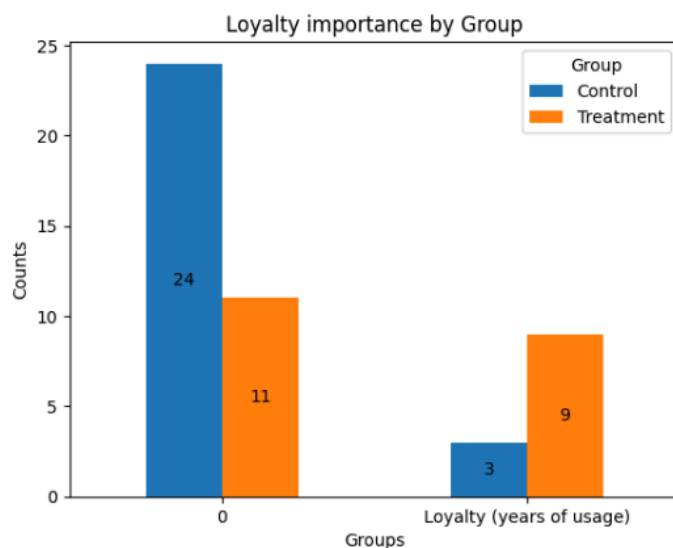
Group	Try	Switch	Change Preference
Control	3	2	1
Treatment	2	3	1

Treatment Recommendation Effect: Apple Music V/s Spotify

Key Takeaway: Most of the survey takers prefer Spotify over Apple Music and were recommended Apple Music as a result. Within the strata, the means didn't vary too much and there were no statistically significant differences between the treatment and the control group for our success metrics.

The most notable difference in means was for the number of people who were willing to try Apple music in the control group which is also reflected in the overall results. There was a difference of almost 42 percentage points between the two groups and this difference was barely statistically insignificant at a 5 % level with a p-value of 0.0076.

Supporting Data Visualization
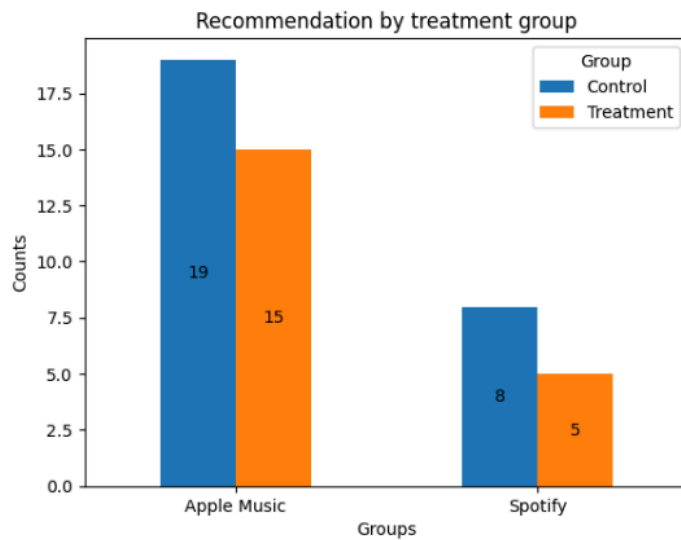
For Apple Music Recommendation

Group	Try	Switch	Change Preference
Control	15	11	6
Treatment	5	4	3

For Spotify Recommendation

Group	Try	Switch	Change Preference
Control	4	2	2
Treatment	3	1	0

Limitations and Future Work

Sources of Bias: The question asking for music streaming service preference pre and post-treatment has a predefined order of listing. Survey takers might feel lazy to reorder the list, which could lead to potential bias in data collection

Ethics: If this experiment design were to be conducted on a larger scale, it's crucial to prioritize ethical considerations and adherence to ethical and legal standards. This includes securing informed consent and safeguarding the privacy and data of participants. The research should clearly communicate any potential risks and advantages, affirm the voluntary nature of participation, and guarantee the confidential treatment of all responses. These steps are essential in respecting participant rights and welfare, while reducing the risk of harm and bias in the study.

Power Considerations: We would like to conduct a power analysis to determine a good sample size that could detect a statistically significant effect.

Sample Size Determination: The sample size we had currently was small and possibly biased, we may want to explore methods to increase the sample size and widen the target audience.

Heterogeneity Assessment: We would like to assess the heterogeneity within the sample population to better understand how behavioral differences may influence outcomes. This may include aspects like music preferences.

Exploring Message Variations: Experiment with different variations of the advertisement message to determine which sets the best for our target audience. We noticed that the message length for the treatment group was much longer than the one for the control group thus making it difficult to draw 1-1 conclusions.

Testing Additional Incentives: Explore the effectiveness of additional incentives beyond a free trial, such as exclusive content access or personalized recommendations.