Project Title: Analysis of Budgeting vs Investment App Popularity in 2023

Team Names: Himaja Jangle, Justin Sterling, Brian Lin, Achintya Kattemalavadi

Date: 12/09/2023

Overview

These days, there are numerous financial apps out on the market, some that focus on investing, some on budgeting, others on saving, and some on all aspects of a person's financial journey. The options for someone looking for a financial app are endless. The problem with going to market with just any financial app is that the market is already so saturated. This generation is learning to be increasingly financially savvy and is looking for direction on how to set themselves up for financial success. With them being on their phones constantly, there is a huge need for mobile tools that provide this guidance. However, there are so many tools already on the market with varying features that have seen varying levels of success with the target market. How can we become leaders in this seemingly saturated market?

The goal of this analysis is to be able to determine if people prefer an app that provides budgeting guidance or investment guidance. Success of an app will be defined by a combination of common app usage metrics such as screen time and churn rate as well as general user feedback. This study will also be diving deeper into if these preferences change based on the group that is using the app. For example, are the preferences of consumers in the age range of 20-30 years different than the preferences of those in the age range of 30-40 years old. Other factors we are looking at are if marriage status or income level changes which type of an app a consumer is more likely to use. In order to study this, it will be important to limit subjects to two apps produced by the same company, with the different purposes of budgeting and investing. Additionally, we will also take a step further and explore how in-app ads affect a consumer's app usage. Half of our subjects will be given the same app with ads while the other half will get the app with no ads. This will also be an important factor the team will be taking into consideration when going to market with their app.

The study will be conducted by providing our sample group with two apps that are developed by the same team, one for budgeting and the other for investment. We will then monitor usage across the time period for the study to glean insights. The data that will be needed to be collected can be split into two categories. Quantitative data will be common usage metrics such as screen time and churn rate while the qualitative data will be interviews that are conducted or surveys sent out to potential consumers to determine which app best fits their needs. From there, we can look to see trends between certain features and popularity, for example, maybe younger people prefer being able to successfully manage their money via budgeting methods than focusing on growing their money via investment methods. This will allow the C-suite team to make a more informed decision on how to best invest their funds.

Intended Audience
C-suite group determining which of their financial mobile apps to invest in further: their budgeting or investment app

Existing Literature

Existing literature suggests that the global financial application market is experiencing rapid growth and features many different types of apps in categories such as banking, digital wallets, and payment systems (link). Some of the largest categories are stock trading apps and budgeting apps (link). In addition, there is research that suggests that smartphone financial applications contribute to improving personal financial capability for people (link). However, one aspect that is lacking in available literature is what types of financial applications users use or seek the most, and what features of these applications draw and sustain larger user numbers. While user reviews of top financial applications provide valuable insights as to what features individual users liked or disliked about different apps, this data does not provide enough controls to serve for formal study.

Anticipated Impact

The goal of this analysis is to be able to determine what type of financial app is going to be the most successful on the market. Ultimately, we are trying to determine if people overall or certain groups of people are prioritizing using mobile apps to budget their money or invest their money. Success of an app in our study will be defined by a combination of general app usage factors as well as survey feedback. Armed with the results of the study, the C-suite team can make a better informed decision on how to best guide their company's priorities.

Research Questions

Main Research Question
Do financial apps geared toward budgeting lead to higher user engagement compared to financial apps geared toward investing?

- Our primary metrics to measure user engagement will be the number of seconds users spend on the app.

Sub-Questions

- Do budgeting apps or investing apps perform better in terms of other success metrics: churn rate, retention rate, active user rate.
- Which type of app leads to users feeling more in control of their finances with a budgeting or investment app?
    - We will be relying on user reviews and survey questions to analyze this subquestion.
- Do different demographics prefer budgeting apps vs. investment apps?
    - The cohorts we are interested in are determined by age group, income level, and marriage status.

Definitions

- Screen Time - Number of seconds participants spend on the app.
- Churn rate - Percent of participants that uninstall the app during the testing period.
- Retention rate - Percent of participants that still have the app. (1 – churn rate)
- Active user rate - Percent of test subjects with the app that use it regularly.
- Budgeting app - Financial application where the primary focus is to help the user manage and save their money.
- Investment app - Financial application where the primary focus is to allow the user to make trades and manage their investments.
- Age groups - We are interested in the following age groups: 21-30, 31-40, 41-50, 51-60
- Income Levels (link) - Lower income (< $47,189), middle income (>= $47,189 and <= $141,568), upper income (> $141,568)

Study Design

Our study will revolve around creating two prototype financial apps, one focused on budgeting and one on investing. We will use stratified sampling to get an equal number of participants in each group, based on age group, income level, and marriage status. Then we will conduct A/B tests on the different outcome metrics between the two apps.

All participants will have the two apps pre-installed on their primary mobile device, but we will not incentivize them to use the apps. At any point during the experiment the participants may uninstall the app as well. The idea behind this approach is to simulate if the two apps were both available on the app market, so we get a more organic view of the two apps being used.

For the qualitative portion of our study, we will analyze both user reviews and conduct user surveys on test subjects' experiences with both the investing and budgeting apps. For user reviews, we will study which types of users were more likely to leave reviews and if these reviews were generally positive or negative. We will also research specific features that reviewers commonly mentioned in their reviews and whether opinions about these features were positive or negative.

We will also send a survey to all users of both apps to ask them about their experiences using the app. The survey will include mostly questions that require the test subject to give their answers on a Likert scale. We will also ask several open-ended questions and again look for common opinions. Open-Ended questions will include:

1. Did you keep the app and continue to use it? Why/Why not?
2. Overall, what did you hope to gain from using this app? Did the app meet these needs?
3. What features did you like/dislike?
4. Was the app easy or difficult to understand? Why?
5. How helpful was the app? Which parts were most helpful?
6. What was the app missing that you would have liked to use?

Data

We will only collect user data from the two apps we release and not from other sources for three reasons. First, we are controlling for many factors in our test to get an unbiased assessment of which app will be more successful in terms of user engagement and will be more

profitable. It would be difficult to control for these factors using data from existing apps. Second, since we will administer these apps, we will be able to collect more information from them than we would have access to from external apps. Third, mixing our app data with external data for testing purposes could violate the assumption of our statistical tests that the data are independent and identically distributed.

We will collect the following quantitative data from users of our budgeting and investing apps:

1. Churn/Retention Rate.
2. Active User Rate.
3. Revenue/Cost/Profit Per User.
4. Total Time spent on each app.
5. Time spent on each feature.

We will collect the following qualitative data from users:

1. User attributes – age, income level, and marriage status.
2. User reviews.
3. User satisfaction surveys

The primary metric we will use to test and compare our apps is screen time. The remaining data will be gathered to further analyze app performance and answer the sub-questions we've listed above.

Sample

Our population of interest in this study is our potential users. We will send out early access invites to try out our apps, with the caveat that the individual would need to install both applications, as well as consent to participate in our study. We will use stratified random sampling, stratifying on the different demographics of interest (age group, income level, and marriage status). If necessary, we will send out multiple rounds of invitations, to ensure that we have at least 100 participants in each group. Our goal here is for each group to be well represented in each study. There will be 24 stratas, as there are 4 categories for age groups, 3 categories for income level, and 2 categories for marriage status. Based on the sampling threshold of 100 for each strata, we will need a total of at least 2400 participants in our study.

Hypotheses

We hypothesize that the investing app will see more success than the budgeting app in our study. This will be reflected in both the quantitative and qualitative data. For example, we hypothesize that the investing app will see higher screen time and lower churn rate. We make this hypothesis for a few reasons. First, in this current day and age, investing is the "trendier" activity while budgeting seems to be more of a chore for many individuals. With investing you

get to see your money grow, while budgeting is just the act of putting restrictions on your spending, which is not as enjoyable. Additionally, budgeting is more accessible without an external tool. For example, anyone can write out a relatively simple budget by hand. However, investing is tougher as you need external information such as share values, value changes over time, and stocks for purchase or sale.

We do anticipate that we are going to see some preference differences in the different groups we are studying. For example, we think that younger individuals may prefer budgeting apps as they don't have enough disposable income to invest and rather just need to make their money last. Overall, we think we would see a bell curve when breaking out age groups and the preference for an investing app over a budgeting app. Younger individuals will likely prefer budgeting, but as they get to middle age, they have more money to spend and can start investing. At the end of the spectrum the elderly may not be investing as much as they are likely pulling their money out of the stock market and spending it. They have less of a reason for it to grow. We also anticipate that there will be a correlation between income and preference for an investing app over a budgeting app. With more income to spend, there is more of a reason to invest extra income than need to budget it. Additionally, we think married individuals will prefer a budgeting app as they have additional folks in the household to support and may potentially have dependents. Therefore, they might have less disposable income to invest.

Variables

Since we will release the apps, we will have access to session data from our chosen cloud provider. This will allow us to collect total screen time from all sessions and calculate our primary variable of interest, total screen time per user. We consider this a more relevant metric for long-term engagement than average session length because it inherently values the number of times users open the app. Overall, we believe using the app many times for short durations and using the app infrequently for long durations should be measured as similar levels of engagement for these particular apps. Additionally, these apps are similar enough that we expect engaged users to spend approximately equivalent amounts of total time on either app.

We will also collect user screen time per feature to give us more detail on user preferences for both apps, and use our performance data to calculate churn/retention rates and active user rates to get a more complete picture of overall engagement. For all metrics, we will look at trends over the testing period to evaluate how user engagement changes over time.

To compliment performance metrics, we will gather user reviews and conduct user surveys on their experiences. As mentioned, we will gather demographic information from our users to evaluate our expectations of user preferences with respect to age, income, and marriage status. We will also summarize the likes and dislikes of users from likert scale answers, and compile a list of the most common replies from open-end question responses. We can use these to further explain our quantitative results and help us make improvements to future versions of the apps.

Though this experiment focuses on user engagement, we would also like to perform a post-experiment profitability analysis, which we clarify here. We first calculate the average active user rate per test subject for each app using retention rates and active user rates. We will then multiply this rate by the average profit per active user to calculate the profit per test subject for each app. Using profit per test subject, we will determine which app type was more profitable during the testing period.

Statistical Methods

We will test our main hypothesis using a paired-sample t-test to compare the sample mean of the total time users spend on the investing app against the mean of the total time users spend on the budgeting app. Our null hypothesis that the two means will be equal will be tested against our alternative hypothesis, that the mean of time spent on the investing app will be greater. We will use a power analysis to determine the appropriate sample size for our test. Additionally, we will run multivariate linear regression to determine if any of our collected user attribute variables explain some of the variance in total screen time and screen time for specific features.

We will use two different methods to study our qualitative data. First, we will use the Wilcoxon Signed-Rank test on each of our Likert scale responses to test which app was more successful regarding that specific question. Second, we will use NLP methods, such as sentiment analysis and topic modeling, to learn more from user reviews and open-ended survey questions.

Potential Risks

Contamination: As we would not be monitoring participants 24/7, there is the possibility that they share information on the applications with other people, and that information eventually makes its way to other participants in the study, influencing their usage habits or opinions on the applications. This includes potential posts on social media. Non-participants of the study may also hear of the trial and introduce new influences on the thoughts of the participants as well as their usage habits.

Falsely Relating Issues: Participants may attribute unrelated issues, such as the performance of their trading portfolio or the state of their bank accounts due to poor spending habits, to their opinion and usage of the applications themselves. This is despite the fact that the applications only provide avenues to manage finances in different capacities and have nothing to do with the actual state of the participants' finances and trading portfolios.

Differing Technology: Different participants will have different smartphone models and will be in areas with different cellular network or wifi coverages and performance. As such, differences in their experiences and opinions with the applications could be due to these external factors. The most that can be done is to ensure that the applications developed are as lightweight and widely compatible as possible, and require the least network communication possible. An idea for a

follow up study would be to see if users are willing to deal with frustrations due to network or phone performance to be able to use certain features.

Length of Study: The amount of time generally required for each application to be useful to the participants may differ. For example, it may be normal that participants manage their trading portfolios every day, but only manage their budgets once a week. While the study would only reflect the results of the usage time and after-study opinions, participants may have actually found the budgeting application just as, if not more useful. However, that would not be reflected by the usage data.

Privacy: Although data will be anonymized, and no information will be collected on the state of each participant's finances, there is always the risk of unintended data leakage to either the researchers running the study or external parties.

Timeline and Deliverables

- *Month 1:* Send invitations to the selected participants with details of participation, terms and conditions, reward details, and the data usage policy. Request information about their age and income as part of the stratification strategy.
- *Month 2:* Finalize the selected sample of participants by checking their properties against predetermined sampling thresholds for different groups (age, income level). Send the selected sample participants consent forms and instructions to get started with the study.
- *Months 3 - 9:* Collect data from the participants in the sample without contacting them.
- *Month 10:* Send out a qualitative assessment survey to the participants and request a response.
- *Month 11:* Begin work on the research report and the presentation based on the findings.
- *Month 12:* Deliver the finalized research report and present the high level findings to the C-suite.

Next Steps

Following our analysis outlined in the current design, there are additional factors that need to be answered to help us create the best financial application on the market. Below are suggested ideas for additional research questions:

- *Do in-app ads deter people from using their apps?*
- *Are people willing to pay extra for premium features like removing ads and educational content? How does this affect the user engagement?*
- *Overall, which type of app is more profitable to release - investing vs budgeting, ads vs no-ads?*