**Initial Plan:**

• Determine the most valuable subscribers

• When does the limited subscriber become more profitable than the lifetime

• The long term limited ones - how many of them are present? How many of them are more profitable than limited subscribers?

• Understanding the subscriber segments present in the database

• Create exploratory visualizations across different variables

• Identify the most likely subscribers who could be sold additional products or services

• We assume that for a customer to keep auto renewal on, they are inclined to continue spending money on their Rosetta Stone subscription.

• So we will model what makes a customer most likely to be an auto-renewer in order to discover their characteristics

• Identify the subscriber profile of those not continuing with their usage of the product and identify the barriers to deeper subscriber engagement where possible

• We will need to find which subscribers have been lost

• Then we’ll create a predictive model for ‘lost’ subscribers

• Outline any business relevant opportunities that are present from your analysis of the data not covered above

• App usage

• Language learning

**Analytical plan:**

Problem 1:

To determine the most profitable customers, we decided to calculate Customer Lifetime Value for limited and lifetime subscribers in USD. We did so by creating a basic linear function, with the initial purchase amount as a constant and the price of a renewal times the customer’s amount of renewals. For yearly subscribers, their median initial purchase was $0 and renewal was $119. For quarterly subscribers, their median initial purchase was $36, with a renewal of $79. In order for either of these subscribers to have a higher CLV than lifetime subscribers, they had to exceed the median CLV of the one-time-paying lifetime subscriber: $199. Looking at the data, the average amount of renewals for quarterly subscribers was 1.32 and 1 for yearly subscribers. On average, the limited subscribers have a lower CLV than the lifetime subscribers. Therefore, Rosetta Stone should focus on attracting more lifetime subscribers until it can raise the amount of renewals of its limited subscribers.

Problem 2:

Knowing that visualization can be an incredibly informative tool, we created multiple plots to look at the different subsets of subscribers. We found that compared to Europe, US/Canada was dominant and many of the subscribers paid in USD. The majority of purchases recorded demonstrated many initial purchases, which also included the lifetime subscriptions. Overall, there are many more limited subscribers, who subscribe for a year or a quarter, than the lifetime subscribers who pay a large sum up front. The options to access Spanish or ‘ALL’ of the languages were the most popular among subscribers, and Spanish was consistently popular throughout the time period. Notably, the start of the COVID-19 pandemic saw an enormous rise in lifetime subscribers buying access to all languages offered.

Problem 3:

For our third objective, identifying those who are likely to be sold extra products, we started with the assumption that those who have auto renewal on are the ones to target. This is because we infer those who are willing to commit early to the payment are more open to potential purchases. To find out who these people are with auto renew on, we created a logistic regression model which predicts whether or not a user will have auto renew on. This model achieved us an accuracy score of 0.77. From this, we received coefficient values which were exponentiated to create an odds coefficient. From these coefficients, we were able to see that users learning the most common languages were the most likely to have auto renew.

However, this was not our first attempt at creating a model to encapsulate the users most likely to purchase extra products. In our first attempt, we created a linear regression model to capture the users who are staying around for a long time. This didn’t work out, as lifetime subscribers threw off the numbers, but were still important to capture here. After battling to try to get the accuracy up, we moved on to the next model.

Problem 4:

For the fourth problem, identifying the profile of those not continuing service , we created a logistic regression model to predict whether a subscriber will be lost. To determine lost subscribers, we labeled them as all of those whose membership is set to expire before March 31, 2020. Once we created the lost subscriber variable, we were able to create the model. With the model, we were able to achieve accuracy of 0.825. After exponentiating the coefficients to get the odds, we found very similar results to the logistic regression model that predicts a user’s auto renew status. This meaning, the most important coefficients were the most popular languages, including English, Italian, French, and Spanish.

For our first attempt at this model, we created another logistic regression model meant to predict whether a subscriber is a limited subscriber. The goal of this model was to see what made up the characteristics of a user who is not committing to a lifetime subscription. This model performed up to par, however, we determined it didn’t tell us what we really need to know. This is because Rosetta Stone is not losing the service of every single limited member.

Problem 5:

For this section, we focused on things we found in the other sections that we weren’t necessarily looking for. One major finding of ours is that those who are using the most common languages are the most likely to move on from Rosetta Stone. This led us to our conclusion that the company needs to focus on encouraging users to learn another language.

Looking at the app data, users mostly just open the app. Most of them are Apple users instead of Android users. Using the app is less popular than using Rosetta Stone on the web, which all suggests that the app needs to be better optimized for consumer use.

We also compared to other competitors through Google trends.

**Appendix:**

Logistic Lasso Regression for Lost Subscriber

Automatic Renewal Logistic Lasso Regression

Created limited subscriber data

Data Cleaning:

• Removed NA’s for App Activity

• Filtered out high purchase amounts that were significantly over the 75th percentile

• Made columns work for regression

Attempted Models:

• Attempted to segment consumers by clustering algorithm- failed