**High level summary of methods:**

The goal of this market segmentation is to develop user segments or "personas" to inform future marketing efforts and product development. This analysis will use an unsupervised machine learning approach to cluster the users into distinct personas. Unsupervised clustering allows for a data driveway to infer structure within the data, aka clusters. The distinct clusters generated can then be interpreted to understand how Duolingo users might be similar or different from each other. Because many of the survey data was collected as categorical variables, this analysis will use a Kmodes clustering algorithm, some numerical data was converted into categories as Kmodes can only handle categorical variables. Kmodes works by iteratively comparing the similarity of each new point k centroids. The new data point is then clustered with the cluster that it is most like, and a new centroid is calculated. The distance between cluster and new point is measured by dissimilarity (total mismatches between data points).

From this clustering algorithm we were able to identify 3 distinct personas from the data.

Chart, radar chart

Description automatically generated**Description of 3 User personals :**

**High Value Customer.**

* Most likely to purchase a subscription.
* Very active with Duolingo app
* Very committed to learning
* High proportion of retirees
* Generally, older (55 – 74).
* Generally, earn more
* Mixed language proficiency

**New Language Students**

* Least likely to purchase a subscription
* Generally younger (18-34)
* Most earn less than 10k
* Learning a language for the first time
* Highest probability of being a student or unemployed

Working Adult Reviewer

* Reviewing a language they have studied before
* Generally middle age (35 – 54)
* Generally earning $26k – $75k
* Most likely to take a placement test
* Highest employment rate

Key Visualizations

Chart, bar chart

Description automatically generatedChart, bar chart

Description automatically generated

Working Adult Reviewer most likely to review an old language

New Language Student most likely to learn a new language

High Value Customers most likely to purchase a subscription

High Value Customers most likely to be very committed

Chart, bar chart

Description automatically generated

Age breakdown of each cluster

Chart, bar chart

Description automatically generated

Recommendations for product changes and marketing campaigns

* High Value Customer.
  + Consider developing a loyalty and referral program targeted for this group. Highlight referral scheme, as word of mouth is the best way to win new customers
  + Have dedicated service representatives if they have issues
* New Language Students
  + Young, group of new language learns. Consider targeting campaigns that will expose them to multiple new languages to help them discover one that interest them.
  + Appeal to young people’s desire to experience new languages with travel marketing campaigns focus on travel
* Working Adult Review
  + Most likely to review an old language, target notifications and marketing campaigns of relearning an old language.
  + Most likely to be working a job, consider sending notifications after work.