Data Analytics Challenge

Customer retention is an important topic for B2C business. Besides making customer happy, we also want to understand what drives retention. As a multiple-product company, we are helping various product and business owners to understand their retention issue specific to their customer profile mix. To give a simulated experience, you are given a made-up data set with the following files:

- quiz_session.csv: There is a quiz flow for potential customer to fill out before they can start using the product. In this file, you will find the answer each prospect provides in the quiz, as well as the time it takes for them to fill out the quiz, along with the device they use in filling out the quiz. Note that not everyone completes the quiz, and for those who do, not everyone becomes a customer. This file only contains the completed quiz, but some did not get converted to customers. Customers are identified by customer_id, and product are identified by product_name. If one customer has completed the quiz flow for two products, there will be two records for this customer in this file.

- customer.csv: This is a dummy data that shows the length of the journey for each customer (measured in number of days), broken out by product. Also, this file has the age information of each customer at the point of sign-up, on top of a binary indicator, indicating if he / she is still a customer at the point this file was generated (let's assume it is today). If he / she is, the journey length is the number of days since the point of the sign-up, and the journey has not completed yet.

Note: customer.csv may contain customers that are not in the quiz_session.csv - could be missing in the collection of quiz data.

Questions (Please use a tool of your choice to write executable code with commentary, accompanied by a brief write-up to illustrate your approach, rationale, and conclusion)

Exploratory warm-up:

1. Which product have the greatest prospect-to-customer ratio?

2. Which product have the greatest challenge with retention in the first 3 months of product use? Do you think the 3 month retention rates vary significantly across different product?

Meat of the problem:

3. We have a repertoire of tactics we can deploy to keep customers happy and engaged, but would like to be smart about who to target with what tactics when. Stakeholders in the customer care and marketing department come to you, asking for guidance. Can you slice and dice the data, with this question in mind, to uncover meaningful insights on the customer retention behavior, and make actionable recommendations accordingly?

Note: for the sake of simplicity, we assume 1 month = 30 days in the context of this problem.