**1. Introduction**

The purpose of this document is to analyze customers’ reviews of Kindle app from Apple App Store and Google Play Store for the past 1 year (5/1/18~5/30/19). Please note that “Review” is different from “Rating”, as in “Rating” is stars (1~5) given by a customer without comments, whereas “Review” requires a customer to enter in their text feedback about the respective app in each platform (iOS or Android), and they tend to have much lower average number of stars given than the average stars given in “Rating”, partly the reason being is more of ‘negative’ perceptions drive more motivation for customers to go out of their ways and leave a review than satisfied customers.

review new and existing sources of demographic data about Kindle customers. It comes out of a request by Jeff W during last year’s OP1 to more deeply explore reading behavior of Kindle customers, and make preliminary recommendations based on the findings.

Today, we are seeking feedback on a few areas:

* Are there notable gaps in the analysis? If so, where?
* What other recommendations would you make based on the data you are about to read?

**2. Data Sources and Methodology**

For this document, we looked at multiple sources of information:

* **Experian:** A random sample of 2% (750K) of U.S. Kindle customers who have reading activity in the past 1 year (total of ~38M). Basic Kindle information – tenure, reading frequency, primary device, multi-device usage, GMS, KU/Prime status, etc. – about each customer was cross-referenced with credit bureau data about income level, age, gender, education, and family. To ensure all the findings in this analysis are representative and equipped with adequate statistical power, all subsets of the sample were diced with various Kindle and demographic attributes with >1000 customers. This information forms the core of the analysis below.
  + Limitations: Among the 750K users inputted into the credit bureau matching tool, we were only able to get a match rate of 59%, thus demographic data are only available for the matched users. This is expected when working with credit bureau data but does introduce sample bias risk. Also, DDS has an anonymization policy that requires any Kindle attributes you send through the tool to apply to a large number of customers to protect anonymity. For example, customers have one of 5 values for Kindle tenure (<1, 1, 2, 3-5, 5-10 years) instead of a specific date. This abides by the policy but decreases the granularity at which we can analyze data.
* **App Annie:** This third-party tool offers basic demographic information about age and gender for each app available in the app store. This data was used to compare Kindle customer trends to similar reading apps on iOS and Android.
  + Limitations: No visibility into the source data, and cross-app indexes are based on the entirety of the app store, not just similar reading apps.
* **Other Internal Analyses:** Several internal teams do regular segmentations of Kindle and Amazon retail customers, as well as customer surveys (Omnibus, new device, etc.). Where applicable, these analyses expand on credit bureau data below.
  + Limitations: Customer segments are not always a perfect match for the data points we think would be most useful for our analysis. Additionally, the previous research may not be available for the same timeframe as internal data.
* **Industry Data**: Additional data points include census data and reports from research centers like Pew that are not specific to Kindle customers or the Kindle apps.
  + Limitations: Results are not easily pivoted to gather deeper insights about sub-customer segments. Additionally, the previous research may not be available for the same timeframe as internal data.

**Other Limitations/Clarifications:**

* **U.S. Only:** All data in this document is for the U.S. marketplace only, as it comprises 54% of monthly active readers. Future explorations may delve into insights for other marketplaces.
* **No Child Accounts:** Data only includes adult accounts, as children are unlikely to have detailed credit histories and third-party tools do not generally provide information for people under 18.
* **Both Shared & Individual Accounts:** The data includes both single-user accounts and multi-user accounts – they are treated the same in the analysis below since there is no easy way to distinguish between them today.

**3. General Demographic Makeup**

Before diving into broader analysis, here are the basic findings about Kindle customers. For a detailed breakdown of these traits, read side-by-side with appendices A, B and C.

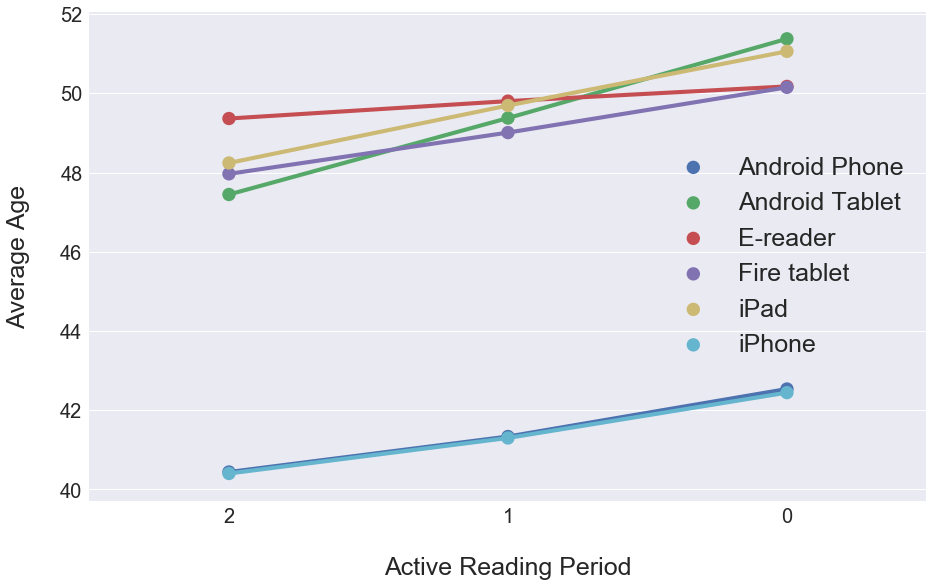
* **Age**: According to credit bureau data, 45% of Kindle customers are over 50. This is consistent with Amazon retail (46% are over age 50), and slightly lower than the 50% of the U.S. population that are older than 50. However, it is significantly higher than the overall population of digital readers. A Pew Research study showed that just 35% of people over age 50 reported reading an e-book in the past year.[[1]](#footnote-2)
* **Gender**: Customers with reading activity in the past year trend heavily female – 61% female, in fact. This is higher than Amazon at large – women comprise 54% of Amazon retail customers – but directionally consistent with readership among the general population. According to Pew, women are more likely to read both print and e-books books than men. 75% of women stated they read a book in the past 12 months (27% read an e-book), as compared to 73% of men (24% read an e-book).[[2]](#footnote-3) However, according to App Annie, Kindle over-indexes on female customers as compared to reading apps like iBooks (47% female) and Play Books (47% female), and even Audible (51% female on iOS, 44% on Android).[[3]](#footnote-4)
* **Household Income:** Kindle over-indexes slightly on high income populations and under-indexes heavily on lower income populations. According to credit bureau data, 33% of active Kindle customers have a household income greater than $100,000 per year, as compared to 28% of the overall U.S. population and 23% of Amazon retail customers. 31% of active Kindle customers have a household income under $50,000 per year, as compared to 43% of the national population and 41% of Amazon retail customers.
* **Education:** Similar to household income, Kindle over-indexes on highly educated customers, and under-indexes on customers with less education. 72% of Kindle customers have some college education or more, higher than the 63% of the general population with the same education level. And while the highest education level for 37% of the U.S. population is a high school diploma, just 28% of Kindle customers are educated at a similar level.
* **Children in Home:** 28% of active readers have children in the home; 72% of active readers do not. This is relatively similar to Amazon at large, where 33% of customers have children in the home, but significantly below the general population average at 38% of households.
* **Race/Ethnicity:** According to Omnibus data, 86% of total Kindle readers are White, 8% are Hispanic, 6% are Black, and 3% are Asian. This maps closely to the overall population of digital readers (85% White), but over-indexes as compared to the U.S. population as a whole, in which 77% of the U.S. population identifies as White. In contrast, Kindle has relatively fewer Hispanic (8%), Black (6%) and Asian (3%) customers than the U.S. population would suggest (18% Hispanic, 13% Black, 6% Asian).[[4]](#footnote-5) At least part of this is likely due to reading population size – according to the Pew Research Center, 27% of White adults, 23% of Black adults and 19% of Hispanic adults have read an e-book in the past 12 months.[[5]](#footnote-6) Since Blacks and Hispanics comprise a smaller overall percentage of the U.S. population, there are fewer potential customers to reach on Kindle. But it’s also a company-wide challenge – Amazon retail is 9% Black and 12% Hispanic, an area that has been identified by that team as a gap to be addressed.[[6]](#footnote-7) And Kindle trends even less diverse than Amazon retail.

**4. Demographic Change Over Time**

Current customer demographics are valuable to examine, but it’s also important to look at how they are changing over time. To measure this, we compared demographic data for customers who were active in the past year (2017-2018) against two similar cohorts from the years before (2015-2016 and 2016-2017), then looked specifically at NTK customers to compare new versus existing customers.[[7]](#footnote-8)

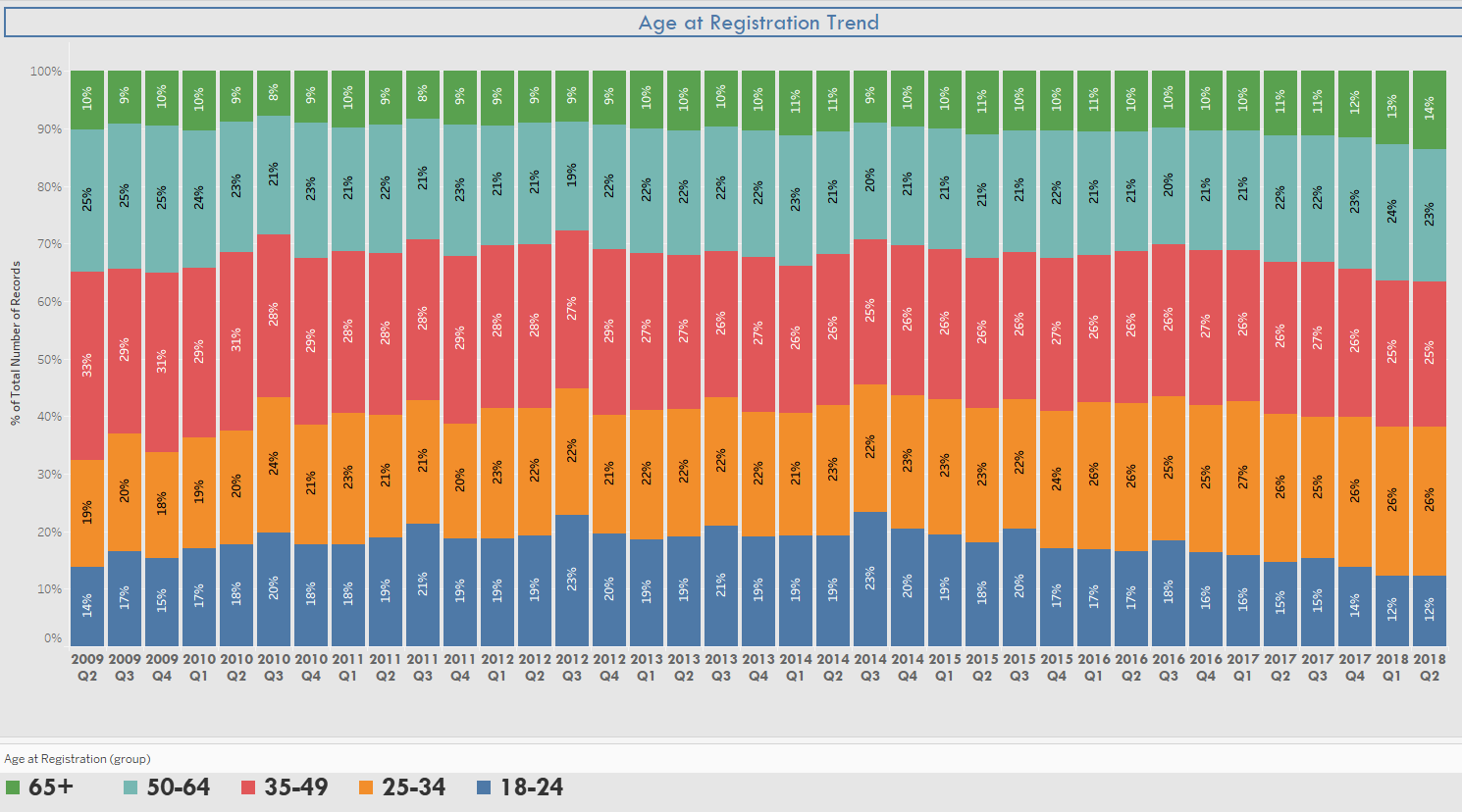
For this, since we do not have a historical snapshot of customer attributes for previous years, we were only able to look at values for which we were able to infer historical records – age since it has a constant rate of change, and gender since it is less likely to change on a broad scale. We were not able to measure changes in other demographic attributes like household income, education, or children because those values change inconsistently over time. If we are interested in diving deeper on those attributes, we can begin tracking now to analyze further in the future.

* **Average age trending up slightly over time:** Kindle readers as a whole trend older than both Amazon retail and the digital reading population as a whole, and have been getting older for the past 3 years in a row. The increase in average age is slightly apparent at the overall level (the average age of active readers from the oldest cohort was 46 years old, and it has since increased to 47.6), without much change in standard deviation, inferring an entire rightward shift.
  + To better understand this shift, we looked at whether the average age changed significantly by primary device used, and saw that the average age is trending up on every major device, although it’s trending up slowest on E-readers (35% of active reading). The average age of customers who read on E-reader is on average 0.8 years older now than they it was in 2015 (from 49.4 years old in 2015 to 50.2 in 2018). In comparison, the average age of iPhone/Android users rose 2 years, and the average age of iPad readers increased 2.8 years (from 48.2 to 51) in the same timeframe.

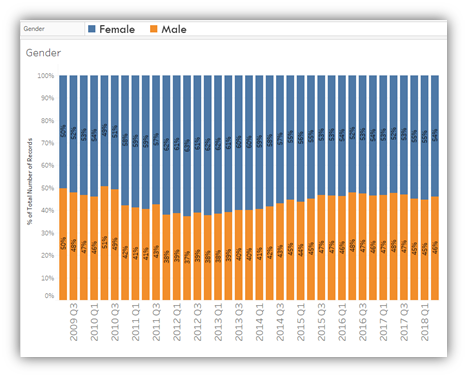
**

*P0 Cohort: 2017-2018 customers, P1 Cohort: 2016-2017 customers, P2 Cohort: 2015-2016 customers*

* **Smaller share of younger customers coming in:** As evidenced by the graph below, the average age of NTK customers trended slightly younger from 2009 to 2016 – from 48 years old in 2009 to 42 in 2016. However, since then it has trended back up (to 43 in 2017), driven both by a larger share of older customers signing up, and fewer young customers doing so. Very young users (18-24) represented about 15% of total monthly NTKs in 2009, which had increased to ~20% by late 2015. However, that group started dropping in early 2016 and was just 12% in the most recent quarter (Q2 2018). This is not just a percentage drop – the absolute number of 18-24 NTK customers is dropping as well. From 2012 to 2015, Kindle was acquiring 800 users per month on average who were 18-24, but from 2016 onward, Kindle has averaged just 560 new users aged 18-24 per month, despite the overall number of NTKs increasing 10% during the same timeframe.
  + Interestingly, we see average age dip (trend younger) every year around August, September, and January. Our hypothesis was that this spike was driven by students signing up to prepare for the new semester of school. We validated this by looking at the primary genre of books those new customers consumed, and noticed they were far more likely to consume textbooks (78% more likely) and comics/graphic novels (38% more likely) than the general NTK population, and less likely to consume other genres like romance (18% less likely).



* **Fewer new women joining:** The majority of active readers are women, and the representation stayed relatively constant over the last few years. Women comprised 62% of active readers in 2015, and comprise 61% in the current year. New customers have also stayed similarly constant – while there was a period in 2012 where women comprised 60%+ of new sign-ups, it has held steady at 55% over the past 3 or so years. Given the gap between customer sign-up rates by gender and the overall gender split, we can infer that women tend to be more engaged and stick around longer than men do.



**5. RFM Segmentation Analysis**

To make some of these demographic insights actionable, we combined customers’ demographic traits with Kindle attributes in an RFM analysis. RFM is a behavioral segmentation of users based on recency (last reading X days ago), frequency (average number of monthly reading sessions), and monetary/spend (average monthly e-book spend) values, providing a lens to determine a high value & engaged customer from low value & disengaged customer. In order to be able to make easier comparisons between customers, we defined “RFM score” by assigning ordinal numeric values with respect to the order (High: 3, Med: 2, Low: 1) based on a customer’s distribution placement in his/her values for each RFM measure, thus the total of 27 (3\*3\*3) segments.

Based on the segmentation, we found that Kindle’s 1 year Active Readers (US) skew towards both extremes, with the largest cohorts of readers (17%) belonging to the low RFM group (R: **L,** F**: L,** M**: L**), and 16% belonging to the group on the opposite end (R: **H,** F**: H,** M**: H**). The next large segment is (R: **M,** F**: L,** M**: L**) at 7%, showing both low frequency and spend with medium recency (last read between 3 weeks ago and 5 months), whom we might consider as lost or on the verge of lapsing depending on various churned reader definitions. Considering more than 80% of Kindle’s ebook revenue comes from less than 20% of users, even though the share of total might be low, LHH (R: **L,** F**: H,** M**: H**) at 0.4%, should be considered as most valuable customers lost or on the verge of losing.

LLH segment, disengaged, but valuable customer in terms of spend makes up 0.7% of the total, and we’ve found that this group of customers highly over-index on primary genre: textbooks (45% of the LLH segment customer’s primary genre), when compared to the base population (7% of all customers’ primary genre is textbooks). See appendix X for full RFM segmentation breakout.

Under the RFM framework, we analyzed the relationships/impacts each of the demographic attribute plays.

**Highlights from this analysis:**

* **Age:** Age is an important factor distinguishing high value customers from low value ones, as older users have higher scores all across recency, frequency, and monetary criteria. Only 4% of younger users (18-24) belong to the highest RFM value segment (High Recency, High Frequency, High Monetary), whereas 12% of older users (50+) belong to the HHH segment, considerably higher than younger users.
* **Gender:** Women are generally more engaged customers than men, although not necessarily more ‘valuable’. Women’s representation along recency and frequency scores slightly over-index men (women are 11% more likely to be in high recency buckets, and 18% more likely to be in a high frequency bucket), but when it comes to monetary score, men and women are essentially the same. 61.1% of men and 61.5% of females belong to the Low monetary segment (score of 1), and 13.4% of men belong to the High monetary segment, whereas 13.6% of women do.
* **Income:** Income is a major driver of recency and frequency, but surprisingly does not seem to be major driver determining to which monetary bucket a customer belongs (High, Medium, and Low). 13% of all users belong to the ‘High Monetary’ category. When adding in an income perspective, however, we found that 12% of low income customers (<$50K/yr) belongs to this group, and 15% of high income users (>$100K/yr) belong to this group
* **Education:** Users with higher education achievement have higher reading recency, read more frequently, and spend more money on e-books than the users with lower education achievement. The higher a user’s education is, the higher their RFM score both at the combined RFM score level and each individual Recency (12% more Bachelor’s+ users’ relative representation in High R bucket than users with less than Bachelor’s degree), Frequency (10% more for Bachelor’s+ than less than Bachelor’s in High F bucket), and Monetary level (34% more for Bachelor’s+ than less than Bachelor’s users). 10% of users who have grad+ education level belong to the highest RFM segment (HHH), whereas 6% of users who have less than HS education belong to the HHH segment. However, education’s impacts on recency and frequency are not nearly as dramatic as they are for monetary value.
* **Primary Genre:** Genre definitely contributes to RFM scores. When a customer primarily reads fiction (mystery, romance, sci-fi, and other fiction excluding international category), their RFM scores rise as compared to non-fiction readers (avg RFM score for fiction primary genre readers is 5, whereas it is 4 for non-fiction primary genre readers). In fact, romance readers lead the pack across the board:
  + Users whose primary genre is romance are most highly indexed among the highest score **recency** segment (42% of them being H recency segment), next in rank are fiction and children’s books primary genre customers. Sci-fi and fiction-other are ranked at the bottom in terms of high recency.
  + Users whose primary genre are romance are most highly indexed among the highest score **frequency** segment (33% of them being H frequency segment), next in rank are users whose primary genre is children’s books. Non-fiction and mystery are ranked at the bottom (read least frequently).
  + Users whose primary genre are romance and textbooks are most highly indexed among the highest score **monetary** segment (17% and 14% belonging to H monetary segment). Users with sci-fi and mystery are ranked at the bottom (spend the least amount of money in e-books).

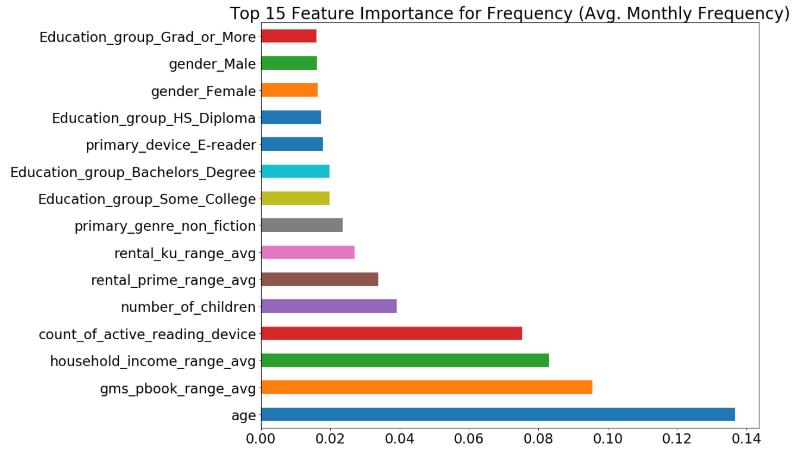
**So what does all this mean? What factors are most important to predict customers with a high RFM score?**

We used decision tree and random forest models to build a model using Kindle attributes and demographics as predictors of RFM. The findings showed that today, 3 demographic attributes show up in the top 5 indicators of customer value:

1. Age
2. Spend on Physical Books
3. Income
4. Count of Active Reading Devices
5. Number of Children

Few other interesting important attributes are primary genre: Non-fiction, and primary device: E-reader, highlighting the fact that depending on the fact whether or not someone has his/**her primary genre as non-fiction**, and **primary device as e-reader** can highly influence (in both directions) that user’s overall value under RFM lens.

For example, here is the relative feature importance for predicting the monetary element of RFM (longer is better):



**6. Other Insights**

After looking at customers overall and by RFM, we looked at a few cross-sections of customers by attributes we thought would be most impactful. These findings did not easily fit into the rest of the analysis, but we thought were valuable to share.

**A. Does primary surface heavily impact customer value and engagement?**

Yes, it does, although it’s more about screen size than overall surface, and age plays a significant role as well. E-reader is the clear leader in customer value; customers who primarily read on that device have the highest overall monthly GMS at $6.62/month. iOS is next at $4.49 per month, closely followed by Fire tablet is next at $4.42 per month in GMS, and trailed by Android at $3.60.

But looking at primary surface by size instead of operating system tells a more nuanced story. It shows that value is more due to screen size than due to iOS users being more valuable as a whole. Android and iOS GMS for phones are similar (and low) – on average, users whose primary surface is Android phone spend $3.36 on e-books per month, whereas iPhone users spend $3.57 – fairly close. And they really deviate on tablet. iPad users spend on average $5.49 per month, whereas Android tablet customers spend $4.23 per month. Due to the caveat with GMS attribution, KCBI team has been using identifying primary device based on total positions consumed at the individual customer level, aggregating GMS at the customer level for the period of interest, and calculating spend per customer metric, and we’ve confirmed with them that under this attribution method, iPhone primary device users spend 7% more than Android phone primary device users, but even this difference disappears when we restrict the customers of interests to only the customers who have only one active reading device, concluding that for customers with only one active reading device, Android phone and iPhone primary device users are spending about the same. For a full breakdown of GMS by primary surface, see appendix C.[[8]](#footnote-9)

Cross-device usage also validates that screen size is a stronger indicator of customer value than simple surface. 68% of customers who use just one device read on a large screen size, but 73% of those who read on multiple devices read primarily on larger screens. We hypothesize that the larger screen size provides a superior reading experience, so customers gravitate more to that as they become more engaged.

Age also plays a clear factor in primary surface – 38% of readers ages 18-34 read on phones, as compared to just 20% of 50+ customers. And while 48% of younger readers read on a tablet-size device (E-Reader, iPad, Fire or Android tablet), a full 71% of those over age 50 do. For a full breakdown of primary device by age, see appendix D.

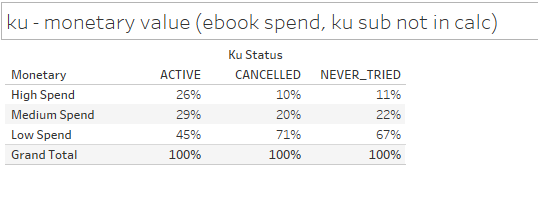
So then are customers on phones lower value/less engaged customers, or are they just younger? Turns out, both are true. Older customers spend and read more, and customers on larger devices spend more. When comparing similar devices, older customers consistently show a higher spend index than younger ones. And when comparing similar ages, larger devices (E-reader, iPad, etc.) consistently show a higher spend index than smaller ones.

*For a further breakdown of spending and engagement by age and device, see appendix E.*

**B. What about KU members?**

*Relevant appendices: G, H*

* **Age:** Approximately 15% of active readers are KU members. Readers trend middle aged to older (29% of active KU subscribers are 35-49 and 46% are 50+, as compared to 26% under 35), but that’s not for lack of trying. Younger users tend to try KU more frequently than older users, but they don’t stick around. 25% of 18-35 year olds have a canceled KU subscription (whereas just 17% of customers 50+ do), but only 12% have an active subscription. 15% of customers 50+ have an active subscription, although interestingly older customers also have the highest percentage of customers who have never tried KU at 68% (see appendix G).
* **Income:** Income doesn’t appear to make a material difference for whether a customer has an active KU membership – 14.4% of customers making less than $50K/year have an active KU membership, as compared to 13.6% of those making $100K+. But income does make a material difference for initial subscriptions and subsequent cancellations. 38% of customers making less than $50K/year have tried KU, whereas just 33% of those making $100K+ have done so. And once they’ve tried it, lower-income customers cancel at a much higher rate (23%) than those making under $100K+ (19%).
* **Gender & Genre:** KU is also slightly more popular with women (65% of KU subscribers), who tend to read romance more than anything else. 37% of reading by women who are active KU subscribers is romance (as compared to just 15% for men), followed by non-fiction (36% of reading by women, in sharp contrast to the 52% by men). Other fiction is relatively similar across the gender split at 13-14% of reading, as are textbooks (7% of women’s reading, 11% for men) and other genres (see appendix H). This romance trend does not carry over for non-KU customers – customers without an active KU membership tend to read non-fiction and fiction far more than romance.
* **Customer Value:** TODO: Add details on customer value



**7. So What?**

All of this information is interesting, but what are the actual lessons learned from this? What kinds of investments should we make in the future based on this data?

* **Focus on retention (Kindle tenure):** Kindle tenure is the single biggest indicator of customer value, above all other demographic indicators. We have many long-tenured customers, but we should ensure we are still getting new ones in the door, and keeping them.
* **Target customers young and old:**  Age is a clear indicator of customer value, and our changing customer demographics have implications for us in the coming years. We should invest in areas like:
  + **Accessibility**: As our population of older customers ages, we should ensure we are appropriately prioritizing accessibility/reading flexibility features. The apps should all work seamlessly in accessibility modes, and we should explore other ways of bundling Audible content or adding TTS when Audible content is not available.
  + **New content types for younger, more diverse customers:** We are already looking into new kinds of navigation patterns for the apps, but there may be additional opportunity in the available content that could help drive up younger NTKs. Given the frequency with which younger customers cancel KU subscriptions, we should look into including content more attractive to younger customers. This may also help make Kindle more likely to be adopted by more diverse customers as well.
* **Raise customer value on phones:** This may appear to be a no-brainer, but it’s worth calling out regardless. Younger customers trend heavily toward phones, but spend less. More lucrative customers trend toward larger devices, but we shouldn’t assume that’s the only way to garner value. Concrete ideas:
  + **Reduce friction:** The forthcoming onboarding project will help introduce new customers to reading on phones, but there’s more to be done around removing the download requirement for new books and making it easier to buy books once you’ve tried them. Competitive analysis of similar reading apps (especially ones that appeal broadly to younger customers, like Libby) may reveal additional areas for investment as well.
  + **Invest in multi-device reading:** If phones won’t go up (or even if they do), then we should get customers reading on both a phone and a larger format device like E-reader or iPad. Improve the transitions between devices, and

The good news is we are already investing in many of the above initiatives over the next few years, and this analysis affirms that these are the right areas to focus on, but we should continue to monitor over time.

**Appendix A: Kindle Reader Segmentation**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | **Age** | | | | **HH Income** | | | | | **Primary Surface** | | | | | |
| **USA (Active Readers past 1 year)** | **Avg.** | **18-24** | **25-34** | **35-49** | **50+** | **<$50K** | **$50K-$75K** | **$75K-$100K** | **$100K-125K** | **$125K+** | **Android Phone** | **iPhone** | **Android Tablet** | **Fire Tablet** | **iPad** | **E-Reader** | |
|  |  | | | | | | | | | | | | | | | | |
| **AMZN Membership** |  | | | | | | | | | | | | | | | | |
| Active Prime Mbr. | **69%** | 65% | 69% | 73% | 68% | 65% | 68% | 71% | 72% | 74% | 68% | 72% | 65% | 68% | 73% | 68% | |
| Active KU Mbr. | **15%** | 13% | 13% | 15% | 15% | 15% | 15% | 14% | 14% | 14% | 17% | 13% | 16% | 15% | 13% | 15% | |
| **Kindle Tenure** |  | | | | | | | | | | | | | | | | |
| <1 year | **6%** | 12% | 7% | 5% | 5% | 7% | 6% | 5% | 5% | 5% | 7 | 10 | 5 | 5 | 5 | 3 | |
| 1~2 years | **6%** | 11% | 8% | 5% | 5% | 7% | 6% | 6% | 5% | 5% | 8 | 8 | 6 | 6 | 5 | 4 | |
| 2~3 years | **7%** | 11% | 8% | 6% | 6% | 8% | 7% | 6% | 6% | 6% | 9 | 7 | 8 | 8 | 5 | 5 | |
| 3~5 years | **18%** | 23% | 20% | 16% | 16% | 19% | 18% | 17% | 17% | 16% | 21 | 17 | 23 | 18 | 15 | 16 | |
| 5+ years | **64%** | 44% | 56% | 67% | 69% | 59% | 65% | 66% | 67% | 68% | 56 | 59 | 58 | 63 | 71 | 72 | |
| **Demographics** |  | | | | | | | | | | | | | | | | |
| **Gender** |  | | | | | | | | | | | | | | | | |
| Male | **38%** | 36% | 38% | 38% | 38% | 35% | 35% | 38% | 39% | 43% | 40 | 36 | 39 | 34 | 40 | 41 | |
| Female | **62%** | 64% | 62% | 62% | 62% | 65% | 65% | 62% | 61% | 57% | 60 | 64 | 61 | 66 | 60 | 59 | |
| **Age** |  | | | | | | | | | | | | | | | | |
| 18-24 | **6%** |  | | | | 8% | 5% | 5% | 5% | 6% | 9 | 9 | 4 | 5 | 4 | 5 | |
| 25-34 | **21%** | 28% | 17% | 18% | 20% | 19% | 29 | 29 | 15 | 17 | 16 | 21 | |
| 35-49 | **28%** | 21% | 28% | 31% | 37% | 29% | 32 | 32 | 26 | 27 | 26 | 24 | |
| 50+ | **45%** | 43% | 50% | 45% | 38% | 47% | 31 | 30 | 55 | 51 | 54 | 51 | |
| **Education** |  | | | | | | | | | | | | | | | | |
| Less than HS | **7%** | 7% | 10% | 6% | 6% | 10% | 6% | 7% | 6% | 4% | 9 | 8 | 7 | 7 | 6 | 6 | |
| HS Diploma | **22%** | 17% | 19% | 15% | 28% | 30% | 25% | 19% | 17% | 12% | 24 | 19 | 26 | 26 | 19 | 18 | |
| Some College | **31%** | 62% | 33% | 33% | 24% | 33% | 32% | 31% | 31% | 25% | 36 | 33 | 32 | 31 | 27 | 27 | |
| Bachelor’s | **26%** | 10% | 22% | 33% | 25% | 18% | 24% | 29% | 31% | 33% | 20 | 26 | 23 | 24 | 29 | 29 | |
| Grad + | **15%** | 4% | 16% | 13% | 17% | 10% | 12% | 15% | 16% | 25% | 11 | 14 | 12 | 12 | 19 | 20 | |
| **HH Income** |  | | | | | | | | | | | | | | | | |
| <$50K | **30%** | 37% | 40% | 23% | 28% |  |  |  |  |  | 37 | 30 | 33 | 32 | 25 | 26 | |
| $50K-$75K | **20%** | 17% | 16% | 20% | 22% |  |  |  |  |  | 21 | 19 | 23 | 22 | 18 | 18 | |
| $75K-$100K | **17%** | 16% | 15% | 19% | 18% |  |  |  |  |  | 16 | 18 | 17 | 18 | 18 | 17 | |
| $100K-125K | **11%** | 10% | 11% | 15% | 9% |  |  |  |  |  | 10 | 12 | 10 | 11 | 12 | 12 | |
| $125K+ | **22%** | 21% | 19% | 22% | 23% |  |  |  |  |  | 16 | 23 | 17 | 17 | 27 | 27 | |
| **Marital Status** |  | | | | | | | | | | | | | | | | |
| Married/Partnered | **79%** | 43% | 58% | 81% | 90% | 61% | 80% | 90% | 86% | 87% | 72 | 74 | 81 | 82 | 82 | 80 | |
| Single | **21%** | 57% | 42% | 19% | 10% | 39% | 20% | 10% | 14% | 13% | 28 | 26 | 19 | 18 | 18 | 20 | |
| **Child in the Home (Y/N)** |  | | | | | | | | | | | | | | | | |
| Y | **28%** | 24% | 24% | 45% | 19% | 19% | 29% | 36% | 33% | 29% | 31 | 30 | 26 | 30 | 26 | 23 | |
| N | **72%** | 76% | 76% | 55% | 81% | 81% | 71% | 64% | 67% | 71% | 69 | 70 | 74 | 70 | 74 | 77 | |

*Highlighted cells indicate variance from Kindle average.*

**Appendix X: RFM Detailed Methodology and Clustering Visualization (U.S. Reader)**

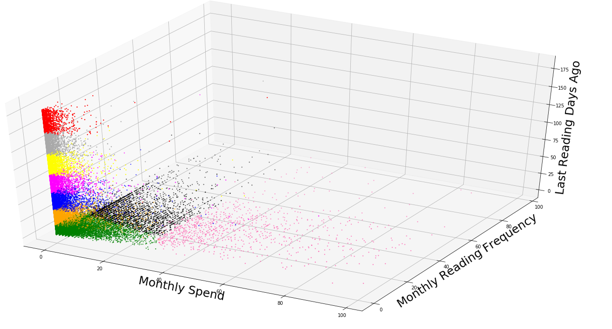
RFM is a behavioral segmentation of users based on recency, frequency, and monetary measures, providing a lens to determine high value, engaged customers from low value, disengaged customers.

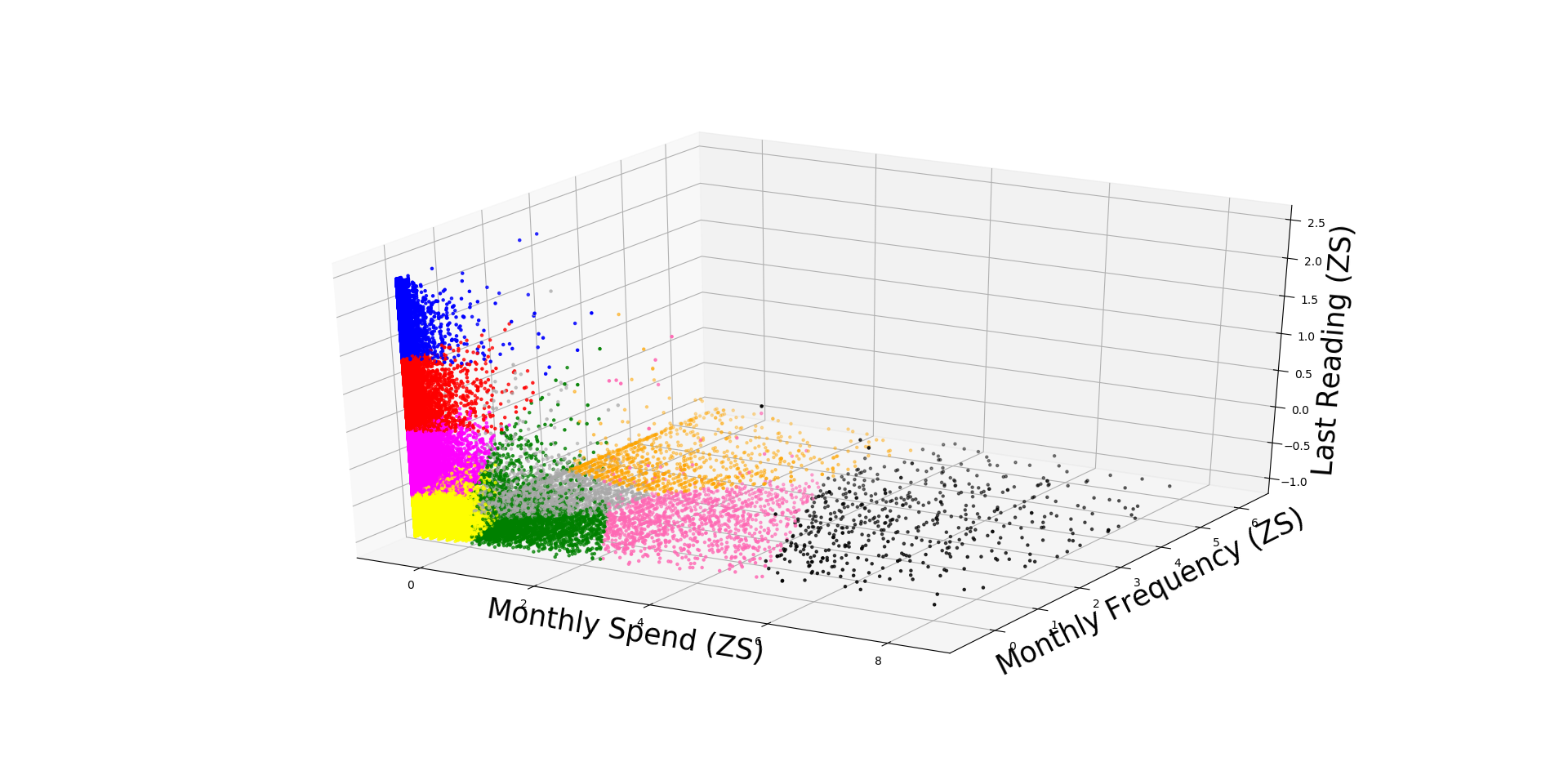
* **Recency**: Based on last reading activity, split into three ordinal values: High:3, Medium:2, Low:1
* **Frequency**: Based on average monthly reading sessions for the past 12 months, split into three ordinal values: High:3, Medium:2, Low:1
* **Monetary**: Based on total e-book purchases (GMS) for the past 12 months, split into three ordinal values: High:3, Medium:2, Low:1

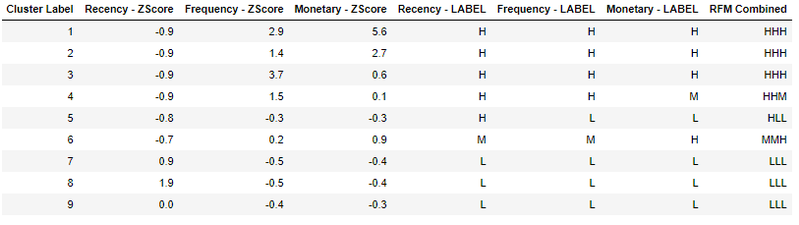
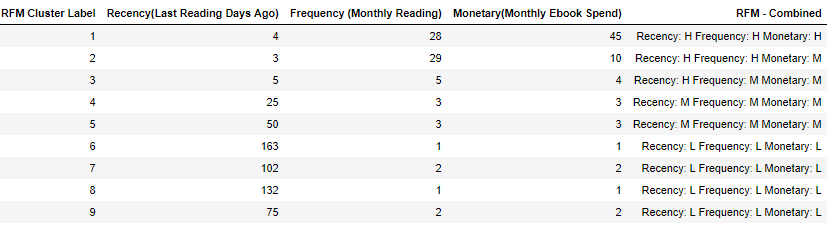
Under RFM lens, the highest value segment: High recency(3), High frequency(3), and High monetary(3) would have a score of 9, and 3 for Low recency, frequency, and monetary, thus RFM score range: 3~9

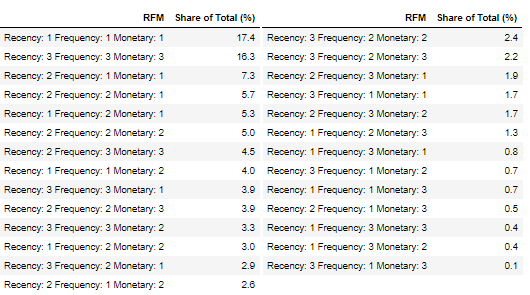
**Clustering Analysis along RFM**

We ran a K-means clustering algorithm to understand representative groups of customers along R,F, and M values, with varying K values (determines how many clusters/groups we want to end up with), and chose K=9 to display here as we believe it captures major cohorts of customers grouped together, which we can check visually from the below chart (each cohort represented by each color).







****

**Appendix B: Gender & Age Breakdown Across Apps**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **iOS** | | | | | **Android** | | | | |
| **App** | % Female | % Male | Age 16-24 | Age 25-44 | Age 45+ | % Female | % Male | Age 16-24 | Age 25-44 | Age 45+ |
| **Kindle** | 59% | 41% | 7% | 50% | 43% | 52% | 48% | 17% | 37% | 46% |
| **iBooks** | 47% | 53% | 13% | 49% | 38% | - | - | - | - | - |
| **Audible** | 51% | 49% | 9% | 58% | 34% | 44% | 56% | 22% | 41% | 37% |
| **Libby** | 52% | 48% | 7% | 53% | 41% | 53% | 47% | 21% | 38% | 41% |
| **Google Play** | - | - | - | - | - | 47% | 53% | 23% | 39% | 38% |
| **Nook** | 59% | 41% | 9% | 47% | 44% | 71% | 29% | 17% | 27% | 56% |

**Appendix C: Kindle Customers by Race & Ethnicity**

|  |  |  |  |
| --- | --- | --- | --- |
| **Breakdown** | **% of U.S. Population** | **% of Digital Readers** | **% of Kindle and/or KU readers** |
| **% White** | 77% | 85% | 86% |
| **% Hispanic** | 18% | 8% | 8% |
| **% Black** | 13% | 7% | 6% |
| **% Asian** | 6% | 4% | 3% |
| **% Other** | 4% | 5% | 4% |

*Data from Omnibus via* [*Kindle Readers by US Race & Ethnicity*](https://share.amazon.com/sites/KindleREAD/_layouts/15/WopiFrame.aspx?sourcedoc=%7b2F3E0284-43A8-4D87-AE15-CEEB0AEEB9BB%7d&file=Kindle%20Readers%20by%20US%20RaceEthnicity%20(Jan%202017%20%96%20Dec%202017).docx&action=default)*, 2018. Numbers do not add to 100% because Hispanic is an ethnicity – Hispanics can be a member of any race. Data cannot be cross-referenced with appendix A due to different sources.*

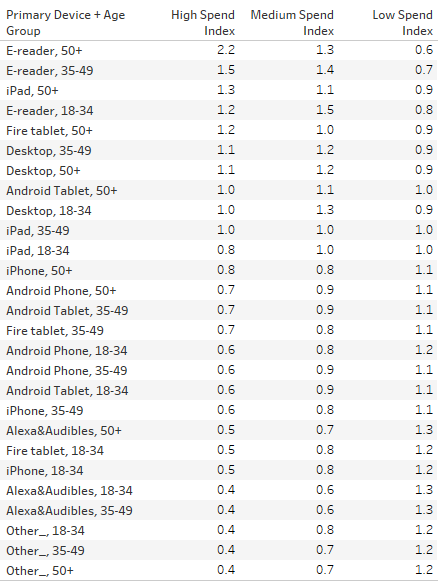
**Appendix D: Kindle Customers by Primary Device and Spending**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Primary Device** | **P0 Mean GMS (monthly Avg. ebook spend)** | | **STD GMS** | |
|  | **All # active devices** | **1 active device** | **All active devices** | **1 active device** |
| **3P Android** | **$3.60** | **$2.48** | **$8.19** | **$5.35** |
| *Android Phone* | *$3.36* | *$2.34* | *$7.38* | *$4.46* |
| *Android Tablet* | *$4.23* | *$2.89* | *$9.96* | *$7.31* |
| **iOS** | **$4.49** | **$2.78** | **$10.13** | **$6.36** |
| *iPhone* | *$3.57 (6.3% more than Android Phone)* | *$2.34 (0% more than Android Phone)* | *$7.95* | *$4.83* |
| *iPad* | *$5.49* | *$3.36* | *$11.98* | *$7.88* |
| **E-Reader** | **$6.62** | **$4.53** | **$12.39** | **$9.02** |
| **Fire Tablet** | **$4.42** | **$2.83** | **$10.15** | **$6.29** |
| **Other** | **$3.59** | **$2.50** | **$8.43** | **$6.04** |

**Appendix E: Kindle Customers by Age and Primary Device**

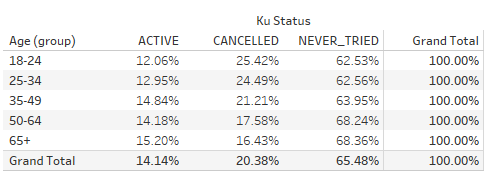
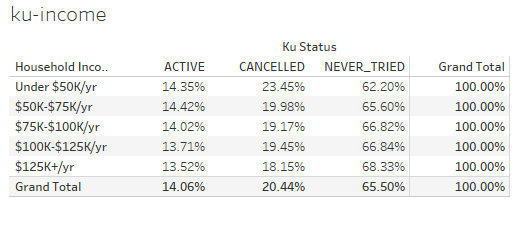
|  |  |  |  |
| --- | --- | --- | --- |
| **Primary Device** | **Age 18-34** | **Age 35-49** | **Age 50+** |
| **Alexa & Audibles** | 2.6% | 3.2% | 2.4% |
| **Android Phone** | 16.5% | 14.5% | 8.6% |
| **Android Tablet** | 3.1% | 4.3% | 5.7% |
| **Desktop** | 6.0% | 3.2% | 3.4% |
| **E-Reader** | 16.6% | 16.0% | 21.1% |
| **Fire Tablet** | 17.9% | 22.6% | 26.2% |
| **iPad** | 10.6% | 13.7% | 17.8% |
| **iPhone** | 21.5% | 18.5% | 10.9% |
| **Other** | 5.3% | 3.9% | 4.0% |

**Appendix F: Kindle Customer Spending Categories by Age and Primary Device**



*Data are ranked by high spend index descending.*

**Appendix G: KU Status by Age & by Income**

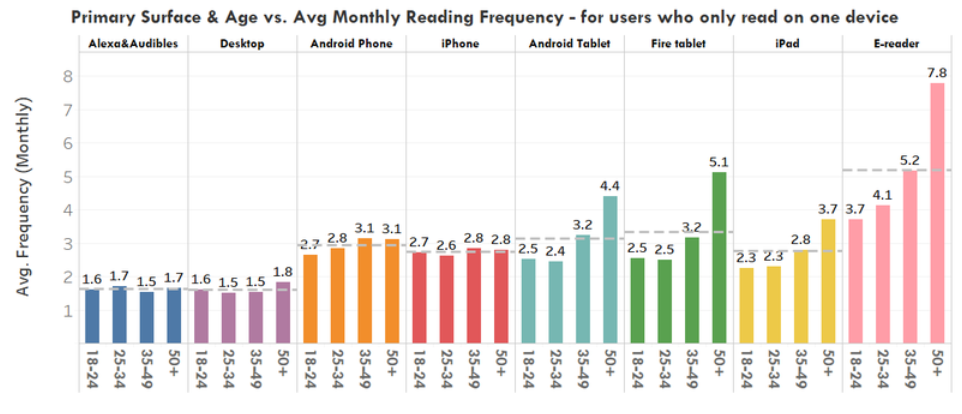
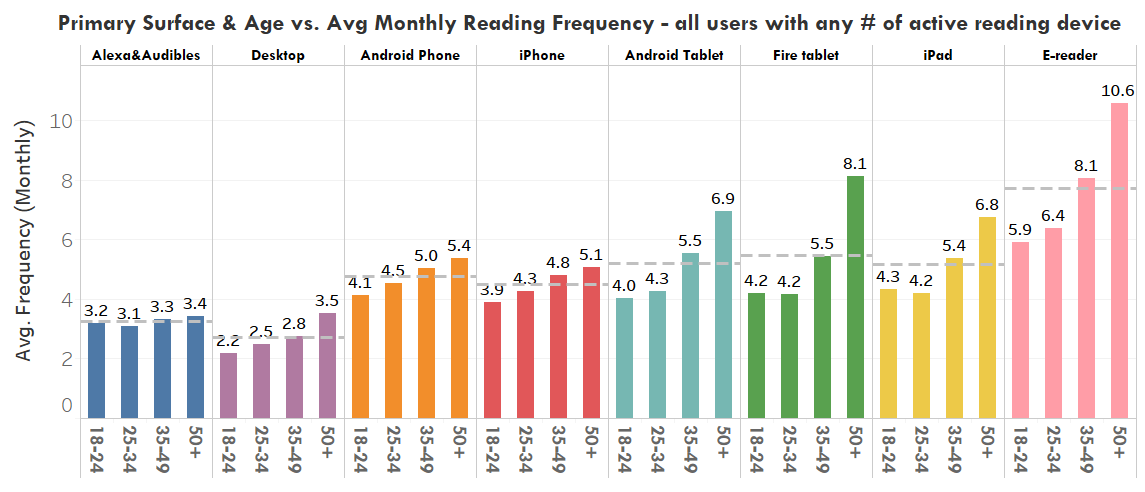
**Appendix H: KU Reading by Genre and Gender**

|  |  |  |  |
| --- | --- | --- | --- |
| **Primary Genre** | **Overall** | **Female** | **Male** |
| Romance | 27% | 37% | 15% |
| Non-Fiction | 43% | 36% | 52% |
| Fiction | 14% | 14% | 13% |
| Textbooks | 8% | 7% | 11% |
| Children’s | 5% | 5% | 5% |
| Mystery | 1% | 1% | 1% |
| Comics & Graphic Novels | 1% | 1% | 2% |
| International | 0% | 0% | 0% |
| Scifi and Fantasy | 0% | 0% | 1% |

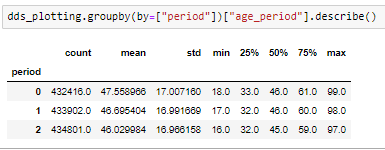
**Appendix G: Amazon Retail Segmentation**

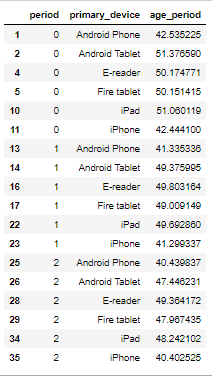
|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| USA | Wtd Avg | 18-24 yo | 25-34 yo | 35-54 yo | 55 + years | <US$ 50K | US$ 50-$US 100K | $US 100K+ |
| % of Population | 33% | 9% | 19% | 35% | 37% | 41% | 35% | 23% |
| % of Spending in Amazon Categories | 15% | 9% | 14% | 17% | 14% | 11% | 15% | 20% |
| Current Paid Prime Member | 39% | 44% | 50% | 41% | 30% | 27% | 42% | 53% |
| Current Free Trial Prime Member | 3% | 9% | 4% | 2% | 2% | 3% | 3% | 2% |
| Demographics | | | | | | | | |
| Female | 54% | 58% | 55% | 54% | 53% | 61% | 53% | 44% |
| 18 to 24 | 9% | 100% | 0% | 0% | 0% | 15% | 5% | 5% |
| 25 to 34 | 19% | 0% | 100% | 0% | 0% | 20% | 20% | 15% |
| 35 to 49 | 26% | 0% | 0% | 76% | 0% | 21% | 28% | 34% |
| 50 and above | 46% | 0% | 0% | 24% | 100% | 44% | 47% | 46% |
| Children | 33% | 13% | 39% | 53% | 16% | 24% | 37% | 43% |
| % with <$25k HH Income | 19% | 41% | 19% | 15% | 18% | 46% | 0% | 0% |
| % with $25k to <$50k HHI | 22% | 28% | 25% | 19% | 22% | 54% | 0% | 0% |
| % with $50k to <$75k HHI | 17% | 11% | 20% | 17% | 18% | 0% | 49% | 0% |
| % with $75k to <$100k HHI | 18% | 7% | 17% | 19% | 19% | 0% | 51% | 0% |
| % with $100k+ HHI | 23% | 13% | 19% | 30% | 23% | 0% | 0% | 100% |
| % 4-Year+ College Graduates | 56% | 31% | 66% | 60% | 52% | 39% | 62% | 77% |
| Urban | 29% | 39% | 39% | 29% | 22% | 34% | 27% | 23% |
| African-American (USA Only) | 9% | 10% | 9% | 11% | 7% | 12% | 7% | 7% |
| Hispanic (USA Only) | 12% | 20% | 17% | 14% | 6% | 13% | 11% | 11% |
| Technology Enthusiasm | 45% | 50% | 53% | 47% | 39% | 42% | 47% | 49% |

*Data from* [*Retail Segmentation U.S. Summary*](https://portal2010.amazon.com/sites/XCM-Insights/_layouts/WordViewer.aspx?id=/sites/XCM-Insights/Prime/Onboarding/2%20PRES_RetailSegmentationUS%20with%20Working%20Names_20160427.docx&Source=https://portal2010.amazon.com/sites/XCM-Insights/Prime/Forms/AllItems.aspx?RootFolder=%2Fsites%2FXCM-Insights%2FPrime%2FOnboarding&FolderCTID=0x012000CD31AD7E3A29EE4D96D49F99099BB4EC&DefaultItemOpen=1&DefaultItemOpen=1)*, 2016.*



Fixed age issue for change from p2 to p0:



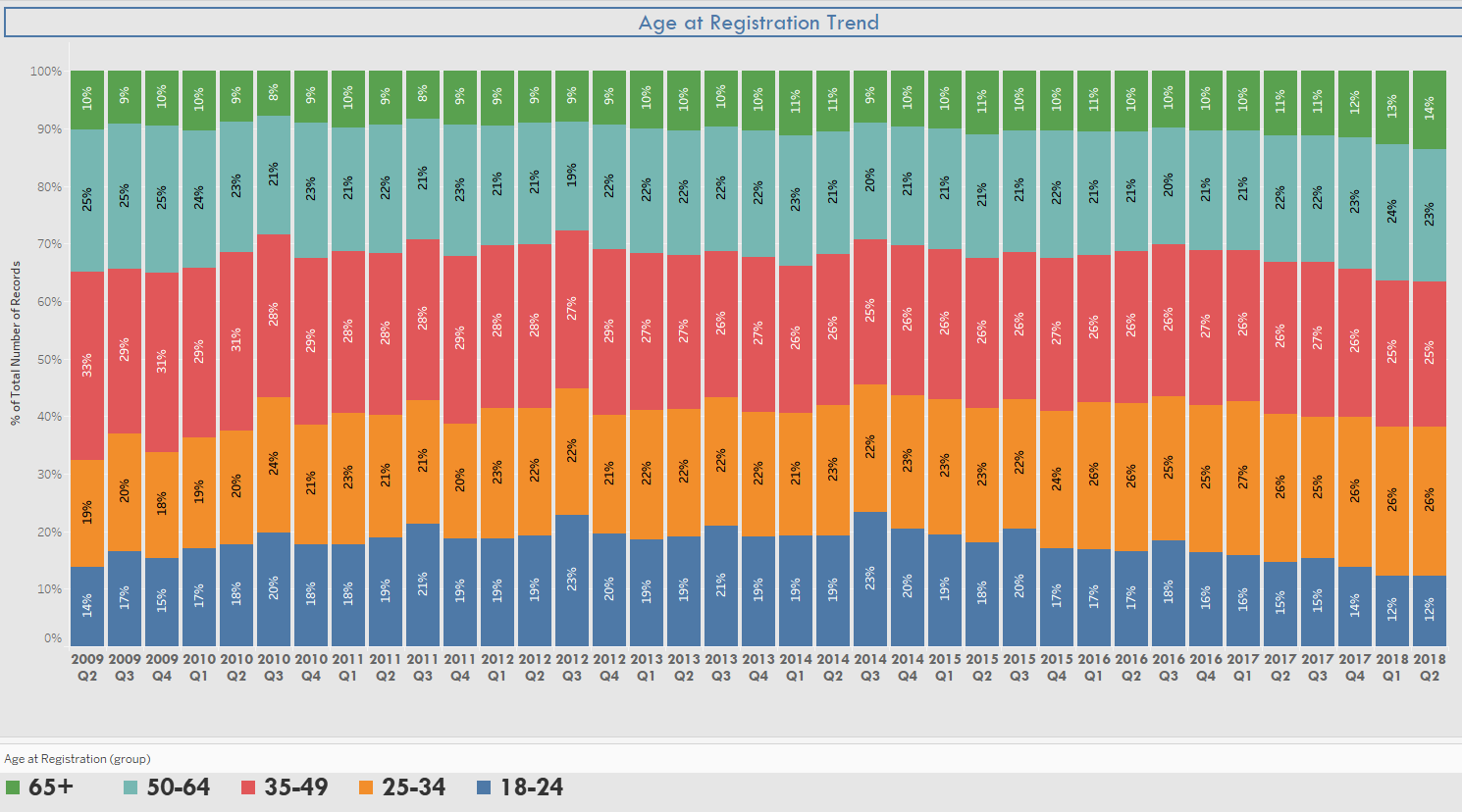


**NOTES – DO NOT PRINT**

**4. New to Kindle (NTK) Customers**

Current customer demographics are valuable to examine, but another important indicator is whether new customer trends are similar to existing customer demographics, and if not, where they differ. To measure this, we looked at changes in demographics for New to Kindle customers since Kindle started. The largest changing trend we identified was in age.

As evidenced by the graph below, the average age of NTK customers trended slightly younger from 2009 to 2017 – from 48 years old in 2009 to 42 in 2016. However, since then has trended back up (43 in 2017), driven both by a larger share of older customers signing up, and fewer new customers doing so. Very young users (18-24) represented about 15% of total monthly NTKs in 2009, which had increased to ~20% by late 2016. However, that group started dropping in early 2017 and was just 12% in the most recent quarter (Q2 2018).



Interestingly, we see average age dip (trend younger) every year around August, September, and January. Our hypothesis was that this spike was driven by students signing up to prepare for the new semester of school. We validated this by looking at the primary genre of books those new customers consumed, and noticed they were far more likely to consume textbooks (78% more likely) and comics/graphic novels (38% more likely) than the general NTK population, and less likely to consume other genres like romance (18% less likely).

Recommendations:

* It may be worthwhile to perform further competitors analysis to understand Libby’s strategy in successfully acquiring younger generation readers in such a short time.
* There seems to be an opportunity to run younger generation focused marketing initiatives or prioritizing CX additions and improvements that address more of the younger customers’ feedback.

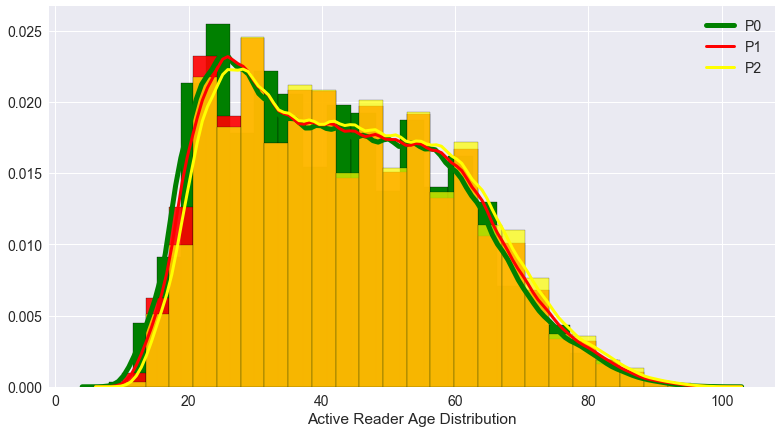
Action Items:

Better understanding of trends in marketing vs organic growth.

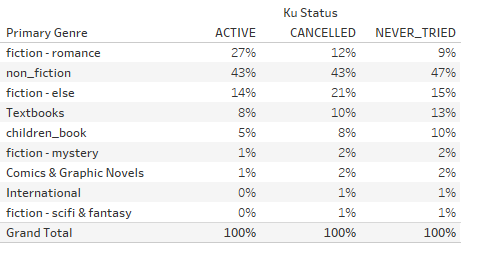
**5. Demographic Changes In Past 3 Years**

To measure how our customer base is changing, we compared demographic data for customers who were active in the past year (2017-2018) against two similar cohorts from the years before (2015-2016 and 2016-2017).[[9]](#footnote-10)

* **Age:** Active readers have been getting younger for the past 3 years in a row. The decrease in average age is slightly apparent at the overall level (the average age of active readers from the oldest cohort was 48 years old, it then increased to 47.6, and is now X), without much change in standard deviation, inferring an entire leftward shift.
  + To better understand this shift, we looked at whether the average age changed significantly by primary device used, and saw that E-reader customers are getting younger at the fastest rate. Customers who primarily use an E-reader are on average 1.2 years younger now than they were in 2015 (from 51.4 y/o in ’15 to 50.2 y/o). iOS customers are also trending younger – the average age dropped 0.5 years each of the years we looked at.



* **Gender:** The majority of active readers are still women, but their representation decreased slightly from 2015-2016 (62% of active readers) to 61% in the current year.



**RFM**

* Although users whose primary genre is children’s book show low recency score, a sizeable chunk of them (18%) fall under high frequency readers, in comparison with only 7% of users whose primary genre is fiction falling under high frequency segment, and 2% of users with non-fiction as their primary genre

**Change**

We’ve looked at the trend over the past three years, three periods in total (P0, P1, and P2), and customers were bucketed into each of these period depending on their reading activity records (customers who were active in multiple periods are duplicated in each respective period they were active in).

Period How it was Defined

P0 Active Readers for the Past 1 year (’17 to ’18)

P1 Active Readers as of 2 Years ago (’16 to ’17)

P2 Active Readers as of 3 Years ago (’16 to ’17)

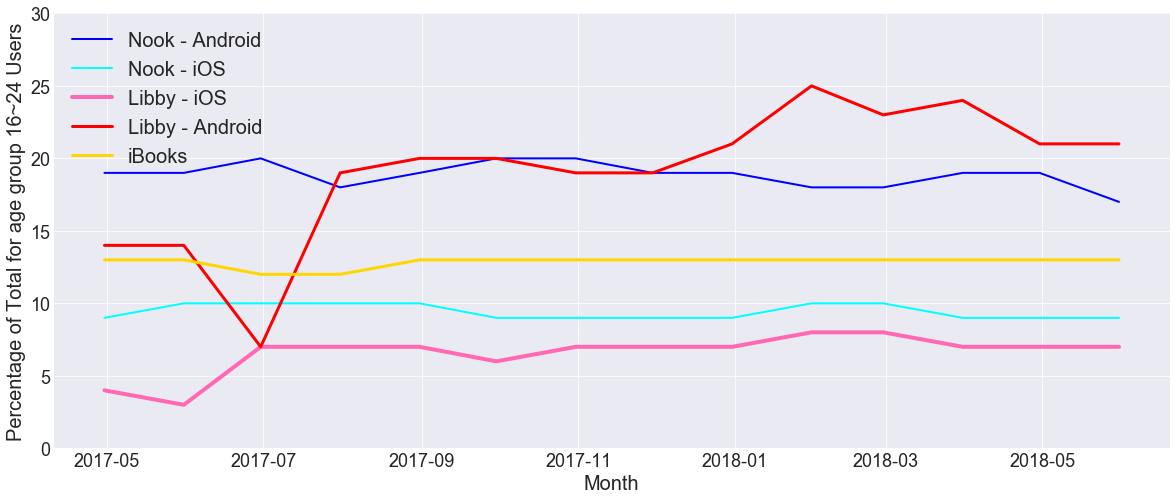
* Which primary surface/device is becoming youngest and oldest?
  + E-reader primary surface active readers are getting younger at the fastest rate. For the past 3 years in a row, mean age decreases by 0.6 years per year. Compared to P2, P0, e-readers primary surface users are on average 1.2 years younger
  + Fire tablet primary surface users are slightly getting older. In P2, average age of active readers on Fire tablet was 49.9 y/o, and in P0 it increases slightly to 50.1 y/o.
    - iOS’s active readers are getting younger faster than Android users. Over the past 3 years, iOS active readers got 0.5 years younger on average, whereas in Android, it didn’t change, with Android user primary surface users base’s age variance increasing at a higher rate than iOS, inferring more diversity in age among active readers using Android device as their primary reading device.

**NTKs**

From 2012 to 2015, Kindle acquired an average of 10,000 new customers per month in non-holiday months. Starting around 2016, Kindle has acquired closer to 11,000 customers per month, an increase of about 10%. With this growth comes changes in the demographic makeup of our customers.

* NTK customers have been becoming younger were getting younger until Q1 ‘17, but becoming older again since then, in contrary to one of our fastest growing competitors, Libby.
  + **Younger user group’s (18~24) representation of the total monthly NTK users increases from ~15% in ‘09 to around 18~20% since then, but starting in Q1 ‘17, it is showing downward trend, and drops to 12% during the most recent quarter (Q2 ’18).** This corresponds to our competitor, Libby’s (iOS & Android) first appearance (Q4 ‘16), and its large popularity growth among younger generation starting in Q1 ’17.
    - **This also corresponds to the downward trend in average age of active Kindle readers from 48 y/o in ’09 to 42 y/o by the end of ’16, trending upwards again starting in ’17 to average of 43 y/o. The trend may imply Kindle slowly losing its relevance among younger customers to our competitors.**
    - Another noticeable trend we spotted from the average age of active Kindle customers chart is the yearly seasonal dips in average age. We found that most of these dips in average age happened in August, September, and January. Our hypothesis to this phenomenon was that students who were preparing for school/semester starts joining Kindle for textbook consumptions are responsible for these dips, and validated this hypothesis by looking at primary genres of the customers who joined when these dips happened, and finding out as compared to the customers who joined in the other months, customers who joined during August, September, and January, had 78% more customers whose primary genre was textbooks, and 38% more whose primary genre was comics/graphic novels, and heavily under-indexed among non-fiction, especially in romance novels (18% less).

*Chart below - Kindle’s major competitors’ % share of total users belonging to younger age group customers:*



**C. Does spending on e-books correlate to spending on physical books?**

Yes. Customers who purchase physical books are more likely to spend highly on e-books as well. 67% of high Kindle spenders have purchased physical books in the last year, as compared to just X% of medium/low Kindle spenders. (TODO – look at primary surface cross-section for this)

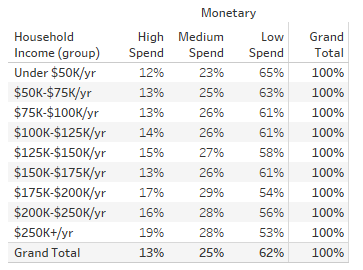
* In addition to Kindle customers trending older than the general e-reading population, App Annie data shows that our population of younger readers is relatively small when compared to other reading apps. Just 7% of Kindle for iOS customers are ages 16-24, as compared to 13% on iBooks. Android customers tend to trend younger than iOS as a whole, with all similar reading apps having at least double-digit percentages of customers in the 16-24 age bracket. But even on Kindle for Android, 17% of readers are ages 16-24, whereas other major apps (Audible, Libby and Google Play) all have more than 20% of customers in the same age bracket. For a detailed breakdown of apps by age, see appendix C.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Breakdown** | **Kindle Customers** | **Amazon Customers** | **U.S. Population** | **E-book Readers\*** |
| 18-24 | 6% | 9% |  | 34% (18-29) |
| 25-34 | 21% | 19% |  | - |
| 35-49 | 28% | 26% |  | 31% (30-49) |
| 50+ | 45% | 46% | 50% | 35% |

*\*Age breakdowns used by Pew do not exactly match internal categories.*

App Annie confirms this gender breakdown for Kindle for iOS, but not across all devices, and this trend does not necessarily persist across other similar apps. Almost 60% of readers on Kindle for iOS are female according to App Annie, whereas just 52% are on Android. Kindle for iOS’ high female ratio is an outlier in the book app space – iBooks is 47% female and Play Books is 53% female on Android. The only other app that is comparable is Nook (59% female on iOS, 71% on Android) The ratios have stayed relatively constant since May 2014, the earliest month for which data is available. For a detailed breakdown by gender, see appendix C.

* Unsurprisingly, customers who make more tend to spend more, although it doesn’t follow a linear trend. Customers who make more than $100,000 over-index on the high-spend end of the RFM analysis, at



Despite the trend for higher education and income among Kindle customers, income does not appear to be a significant indicator of customer value. When looking at income from a recency, frequency and monetary lens (RFM, see detailed methodology in appendix A), customer value is relatively constant across income brackets. 19% of users with an income of less than $50,000 per year are ‘high’ spenders and 56% are low spenders based on GMS for the past 12 months, as compared to users with higher income ($250K+/yr), a group made up of 25% high spenders and 46% low spenders. While there is a gap between 19% high spenders on a lower income and 25% high spenders on a higher income, the 6% gap is significantly less than one might expect given the significant discrepancy in overall income.

* For users with $100K-$125K/yr income, 20% of them are high spenders, and 53% of them are low spenders, not very too much different from low income users

1. http://www.pewresearch.org/fact-tank/2018/03/08/nearly-one-in-five-americans-now-listen-to-audiobooks/ [↑](#footnote-ref-2)
2. http://www.pewresearch.org/fact-tank/2018/03/08/nearly-one-in-five-americans-now-listen-to-audiobooks/ [↑](#footnote-ref-3)
3. See appendix B for detailed breakdown of gender breakdown across apps from App Annie. [↑](#footnote-ref-4)
4. Kindle Content Readers by US Race/Ethnicity, 2018 [↑](#footnote-ref-5)
5. http://www.pewresearch.org/fact-tank/2018/03/08/nearly-one-in-five-americans-now-listen-to-audiobooks/ [↑](#footnote-ref-6)
6. Millennials Diversity Barriers Exploratory Research, Oct. 2017 [↑](#footnote-ref-7)
7. Customers who were active in multiple periods are counted in each. [↑](#footnote-ref-8)
8. TODO: Add context that explains why this number is so different than what we’ve looked at in the past. [↑](#footnote-ref-9)
9. Customers who were active in multiple periods are counted in each. [↑](#footnote-ref-10)