Overall Analysis:

|  |  |  |  |
| --- | --- | --- | --- |
| Found groups: | Described as: | Possible indicators in our dataset | Discovery / diversity Tilt score |
| Active Curator | Regularly curates and listens. | Measure total usage (days & number minutes) | HHH |
| Addict | Lots of time searching, few songs multiple times | Low unique songs, high listening time. | LLH |
| Guided Listener | Uses new songs to generate leads, chooses to engage algorithm | Change in intervention rate (might be difficult to find) | LHL |
| Music Epicurean | Listens to whole album, searches for specific songs. | NA | HLH |
| Music Recluse | Little interaction on platform. | NA | HLL |
| Non-Believer | Self-determines music, doesn’t engage with platform. | Low intervention rate? | LHH |
| Wanderer | ‘Grazer’ in listening, spends time browsing / discovery is goal | High intervention rate. | HHL |

^proportions for each group identified in YouTube and Spotify users.

Known Industry Metrics (Spotify’s 3 streaming habits):

* Discovery : how much new music is sought out
  + Difficult to discover in our dataset, perhaps changing the song as a metric.
* Diversity : range of music
  + High unique number of songs might be a weak proxy.
* Tilt : how actively people curate their streaming
  + Can be easily determined, need to set cut-offs.

HHH (active curator): searches for music, large range, engages curation

HHL (wanderer): searches for music, large range, little curation

HLL (Recluse): searches for music, small range, no curation engagement

HLH (Epicureans): Searches, small range, highly curated  
LHH (non-believer): low search, high range, some curation

LHL (Guided listener): low search desire, high range, low curation

LLH (Addict): low search, low range, highly curated (specific songs)

LLL (NULL)

(see below for analysis)

Predictions:

How do measure search:

* Can use early skipping as a proxy
* Or they have low beginning skips because they pick things?

How to measure range:

* Number of completed songs
* High number of unique songs

How to measure curation / tilt:

* engagement with skipping (play length)

Defining average skip time:

|  |  |  |
| --- | --- | --- |
| Variable | Definition | Suggested average |
| Num\_25 | < 25% | 0.1 (want to weight downward) |
| Num\_50 | 25-50% | 0.375 |
| Num\_75 | 50-75% | 0.625 |
| Num\_985 | 75-98% | 0.86.75 |
| Num\_100 | 98-100% | 1 |

Average time at skip = weighted average of columns.

Deviation of skip = tough?

Percent early skips = Num\_25 + Num\_50 / total skips

Engagement = time per engagement?

Range\_proxy = unique songs / early skips

Search\_proxy = early skips / time (lots of skips for the amount listened)

Tilt\_proxy = total skips (same as engagement?)

Business insights:

* Segmenting users is only useful in how it can be used to increase focus on or engagement with those specific users. This is good in our case, if we can identify groups (of any type) we can justify building a case around them.
* KKBOX has several areas they use to stand out
  + Large Library of South Asian (mainly Chinese) music
  + Lyrics for most songs
  + Innovative “listen with” feature, allowing users to sync their playlists
  + Manual playlist curation.

Questions:

* Users seem to have strong patterns, so each day “should” look like any other for a specific user. This would be good to validate.
* Collecting more data around how users construct and interact with playlists would be useful and might help predict churn. Engaged users “tend” to stay.

Sources and comments in order of relevance:

https://students.washington.edu/yeaseul1/paper/article3.pdf

- [source of model: 18 : https://archives.ismir.net/ismir2015/paper/000012.pdf]

- classify persona's and give examples

- provide data on population level for each type

- provide Spotify and YouTube proportionality data.

? data on playlists may be a strong driver of interest in this platform - especially with the 'listen with' feature.

? do people curate and listen to 1 playlist together?

https://www.kkbox.com/about/en/about

- not available in the US

- "listen with" feature

- has playlist features

- lyrics are a big driver, what kind of user is this?

- Do users combine playlists, lyrics (karaoke), and co-listening? Would this be a good pandemic pivot?

https://www.theverge.com/2017/7/11/15953012/spotify-marketing-data-listening-habits

- sub-categories of below ("easy goers" = background music, "eclectics" = no specific taste)

- (https://spotifyforbrands.com/us/feature/streaming-habits/)

- definitely possible to sell at a particular consumer

- from search (https://www.thedrum.com/news/2017/06/13/understanding-people-through-music)

-- Spotify uses 3 streaming habits

--- discovery : how much new music is sought out

--- diversity : range of music

--- Tilt : how actively people curate their streaming

https://openviewpartners.com/blog/customer-segmentation/#.X5ipI1NKjUr

- A priori seg : based on available characteristics (company size)

- Needs based seg : what customer wants

- Value based seg : by economic value

- need to have a basal hypothesis and test against it

- many variables to segment are only valuable in context (how they can be used in business)

- focusing on 1 or several groups is the way to grow.

https://www.sciencedirect.com/science/article/abs/pii/S0305048300000426

- [seg of online services using clustering]

- fuzzy clustering : points can belong to multiple clusters (useful if we go to characterization)

- paper analyzes music service not users

https://www.midiaresearch.com/blog/music-consumer-segmentation-from-lagging-indicators-to-leading-indicators

- Music aficionados : core customer

-- Spend $5.8 / month

-- only 17% of customers

- forgotten fans

-- 30% of customers

-- above average listening, low spending.

? is there price segmentation in the data? Different $/time for users?

https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.1040.6749&rep=rep1&type=pdf

-[understanding user behavior]

- users have strong patterns for usage

? can we confirm strong usage patterns?

-you can model user session 'stickiness' but that’s not useful for us.

Known Industry Metrics (Spotify’s 3 streaming habits):

* Discovery : how much new music is sought out
  + Difficult to discover in our dataset, perhaps changing the song as a metric.
* Diversity : range of music
  + High unique number of songs might be a weak proxy.
* Tilt : how actively people curate their streaming
  + Can be easily determined, need to set cut-offs.

Deeper analysis of personas:

Active curator:

* Engagement with curation (playlists)
* High discovery
* High diversity
* High tilt

Guided listener

* Wants to hand over control
* low discovery (low search)
* High Diversity (low preferences)
* Low tilt (little interaction)

**Music Epicureans:**

* Music snob, heavy search function and engage in specific songs
* High discovery
* Low diversity
* high Tilt (make choices)

**Music Recluse**

* Private listener, sporadic habits, little known
* High search (specific tastes)
* Lwo range (specific tastes)
* Low engagement (low envestment)

**Non-beliver**

* Low algorithm engagement
* low search
* high range
* high curation But they make choices (weak)

Wanderer

* Enjoys discovery
* High Discovery
* High Range
* Low tilt?/

Addict

* Fixates on specific songs
* Low discovery
* Low diversity
* High tilt (chooses same song repeatedly, low engagement)

Discovery / diversity / tilt

Searching / range / engagement

HHH (active curator): searches for music, large range, engages curation

HHL (wanderer): searches for music, large range, little curation

HLL (Recluse): searches for music, small range, no curation engagement

HLH (Epicureans): Searches, small range, highly curated  
LHH (non-believer): low search, high range, some curation

LHL (Guided listener): low search desire, high range, low curation

LLH (Addict): low search, low range, highly curated (specific songs)

LLL (NULL)

Predictions:

How do measure search:

* Can use skipping as a proxy
* Or they have low beginning skips because they pick things?

How to measure range:

* Number of completed songs
* High number of unique songs

How to measure curation / tilt:

* engagement with skipping (play length)