NetEase Cloud Music Data

Dennis J. Zhang[1], Ming Hu[2], Xiaofei Liu[3], Yuxiang Wu[3], Yong Li[3]
1. Olin Business School, Washington University in St. Louis, St. Louis, MO, USA
2. Rotman School of Management, University of Toronto, Tornoto, ON, Canada
3. Netease Cloud Music Inc., Hangzhou, China

This paper describes the impression/display data and corresponding user, creator, and music content card data from NetEase Cloud Music. This data set is collectively supplied by the Revenue Management and Pricing (RMP) Section of INFORMS and NetEase Cloud Music to support data-driven research in Operations Management. The data contains more than 57 million impressions/displays of music content cards recommended to a random sample of $2,085,533$ users from November 1st, 2019 to November 30th, 2019. For each impression, the data provides the corresponding user activities, such as clicks, likes, and follows. Moreover, the data set also contains information on each user, each content creator, and each content in the impression sample.

Key words: Data Competition, Platform Operations, Revenue Management.

1. Introduction

With the development of faster internet speed and better mobile connections, many people are now streaming music through services such as Spotify and Apple Music instead of purchasing the hard-copy music CDs. In fact, a recent report from the Recording Industry Association of America (RIAA) reveals that streaming accounts for 80 percent of the US music market in 2019, compared with 7 percent in 2010.[1] According to the same report from RIAA, the number of music streaming subscribers in the US rose from 1.5 million to around 61 million from 2010 to the first half of 2019.

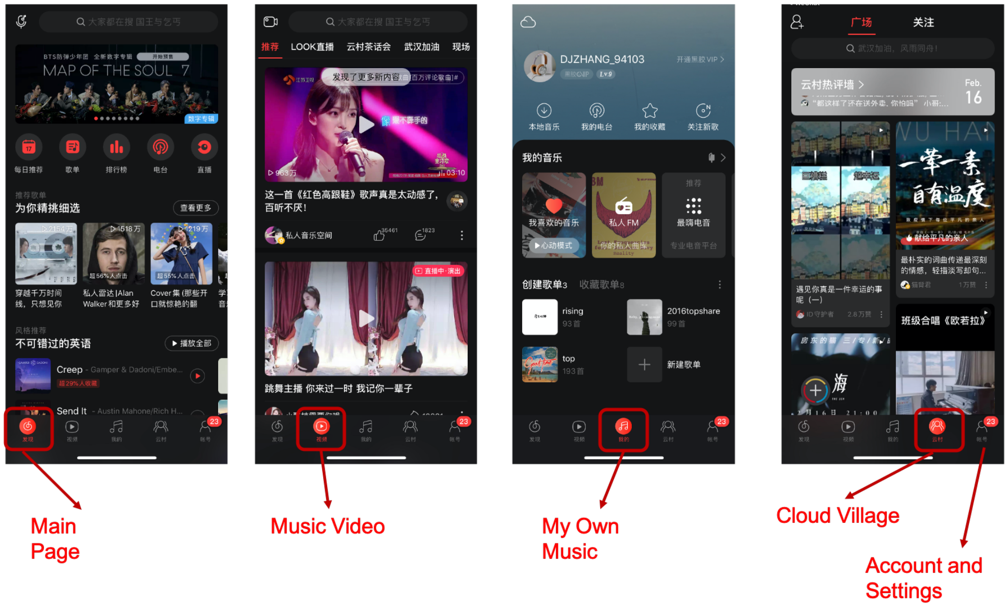
The US is not the only country in which music streaming is reshaping the music entertainment industry. Other countries also observe a similar trend. In this paper, we describe a data set from one of the largest music streaming companies in China—the NetEase Cloud Music (hereafter referred to as NCM). NCM is a free music streaming service developed and owned by NetEase, Inc. It was first launched on April 23, 2013, and then became immensely popular in China. According to a recent report, the music streaming service has around 800 million users in 2019, with a valuation of around 9 billion dollars.[2]

The major product of NCM is its music app (hereafter referred to as "the music app"). Figure 1 shows the main user interfaces of this app. As we can see, there are five main tabs at the bottom

[1] https://variety.com/2019/biz/news/music-streaming-soared-2010s-decade-riaa-1203454233/

[2] https://www.musicbusinessworldwide.com/alibaba-is-spending-2bn-to-acquire-20-of-netease-cloud-music-say-sources/

Figure 1 Five Main Tabs on NCM App



in this app. From the left to the right, the first is the "main page" tab, which consists of the recommended albums and podcasts. The second is the "music video" tab, which contains a single feed of music videos. The third tab in the middle is the "my own music" tab, which shows one's own locally stored music. The fourth tab is the "cloud village" tab, which contains two feeds of short music content cards (hereafter referred to as cards) that are recommended to a user. A music content card can be either a short video with background music or a set of pictures and texts with background music. The last tab is the "account and settings" tab, where users can change their account settings.

In this data set, we will provide more than 57 million impression-level data in the "cloud village" tab associated with $2,085,533$ users in a 1-month-long sample period from November 1st, 2019 to November 30th, 2019. Impression is a commonly used term in the advertising literature, which refers to the display of an advertisement on a web page to a user.[3] In our context, each impression corresponds to a card displayed to a given user on his feed in the "cloud village" through the recommender system. We will also provide users', creators', and cards' characteristic information regarding all users, creators, and cards that appear in the impression data. Our data can be divided into six different tables, and we will describe each table in detail in Section 2.

To help researchers identify practical problems that are of interest to NCM, we discussed with the management group of NCM and provide the following list of research questions:

1. The company defines a user to be inactive if she has a zero or very low average click probability on recommended cards. The company wants to design recommender systems to make inactive

[3] https://www.investopedia.com/terms/i/impression.asp

users active and to make active users remain active. How could the company design different recommender systems to serve users with different activity levels?

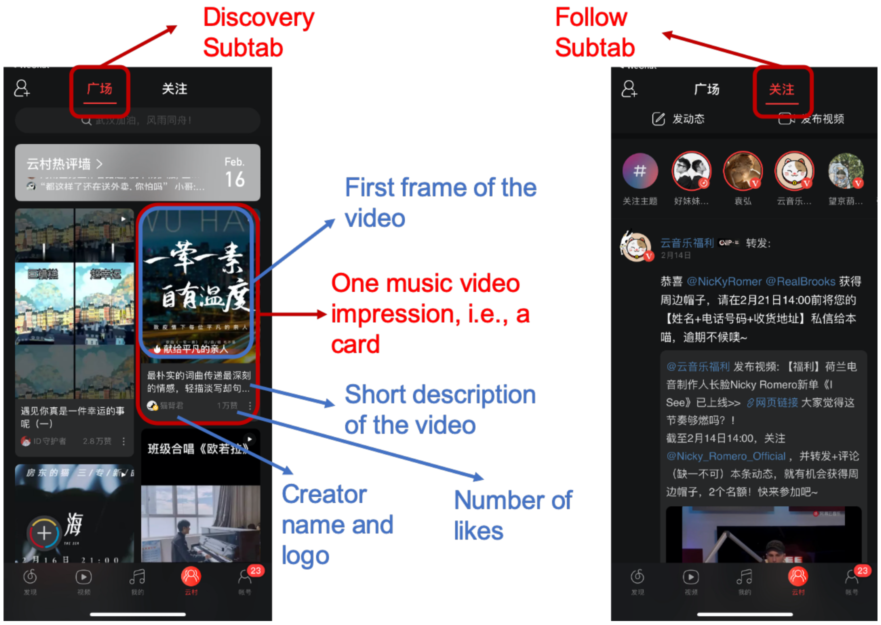
2. What are the characteristics and preferences of active users on the platform? How could the platform predict whether a user will be active or inactive from her early-on actions, such as clicks, likes, and shares? How could the platform design the recommender system to maximize the number of active users?
3. How do different types of feedback information, such as the number of likes and follows, change a creator's motivation to publish new content? How could the platform design the recommender system to encourage creators to create more content?
4. The company's long-term goal on the "cloud village" tab is to maximize the daily number of clicks/plays and the daily number of content created. How could the company create a recommender system that trades off short-term goals, such as the number of clicks/plays in one day, with this long-term goal?

2. Data Description

In this section, we describe all tables in the data set provided to researchers. To ensure confidentiality and user privacy, certain characteristics and key identification information, such as username, user ID, and music genre, are anonymized or dropped. The data set is centered around $2,085,533$ users who have participated in the "cloud village" tab, as shown in Figure 1, during a 1-month-long sample period from November 1st, 2019 to November 30th, 2019. As mentioned above, the "cloud village" tab is one of the five main tabs in the NCM App, and it is a platform where users can post short videos or sets of images with specific music. The users in our data do not contain the entire population of users who access the "cloud village" tab during our sample period. Rather, due to confidentiality, we randomly sampled a subset of users from all who have accessed the focal tab at least once during the sample period.

Figure 2 shows the details of our focal tab, the "cloud village" tab. As shown, this tab can be divided into two subtabs: the discovery subtab and the follow subtab, which are shown on the top of the app. The discovery subtab shows two vertical streams of music video cards that are recommended to a user by the recommender system developed by NCM. Each card contains the first frame of the video or the first picture in a set of images, the creator's information, a short description, and the number of likes that the card has accumulated until the current impression. The follow subtab shows the cards from creators that a user has followed, and these cards are ranked chronologically based on their publication time. Since the majority of the activities in the "cloud village" tab happens in the discovery subtab, and we would like to provide data that may be useful for operations management researchers to test various recommender systems, we focus on the impression data in the discovery subtab.

Figure 2 The "Cloud Village" Tab's User Interface



"都这样了还在送外卖, 你怕吗" 小哥:...

最朴实的词曲传递最深刻的情感，轻描淡写却句...

猫背君 1万赞

遇见你真是一件幸运的事呢（一）

ID守护者 2.8万赞

云音乐福利 转发：

恭喜 @NicKyRomer @RealBrooks 获得周边帽子，请在2月21日14:00前将您的【姓名+电话号码+收货地址】私信给本喵，逾期不候噢~

@云音乐福利 发布视频：【福利】荷兰电音制作人长脸Nicky Romero新单《I See》已上线>> 网页链接 大家觉得这节奏够燃吗？！
截至2月14日14:00，关注 @Nicky_Romero_Official，并转发+评论（缺一不可）本条动态，就有机会获得周边帽子，2个名额！快来参加吧~

Table 1 Summary of Tables in the Data Set

File	Category	Section in the Paper	Data Level	Number of Observations
impression_data.csv	Impression	2.1	impression-level	57,750,395
mlog_demographics.csv	Card	2.2	card-level	252,955
mlog_stats.csv	Card	2.2	card-day-level	4,191,677
creator_demographics.csv	Creator	2.3	creator-level	90,534
creator_stats.csv	Creator	2.3	creator-day-level	2,572,512
user_demographics.csv	User	2.4	user-level	2,085,533

This impression data starts with a data table containing the $57,750,395$ card impressions displayed to users in the discovery subtab. Each impression represents a card shown to a specific user at a particular time during the sample period. Since each impression consists of a user, a creator, and a card, in the other five tables, we will also provide daily information with respect to each user, each creator, and each card that appear in the impression data table. Table 1 provides a summary of all six tables in the data set. In the following subsection, we will first discuss the impression-level table and then move on to the card, creator, and user data tables.

2.1. Impression Data

Table 2 describes each data field in the impression-level data table. This data table contains $57,750,395$ impression-level data points covering $2,085,533$ unique users during the 30-day-long sample period. The table contains 13 data fields. The *userId* data field uniquely identifies each user in the entire data set, and it can be used to join this table with user tables in Section 2.4. The *mlogId* data field uniquely identifies each card and can be used to join this impression table with card tables in Section 2.2. The *impressTime* data field records the epoch time when the impression

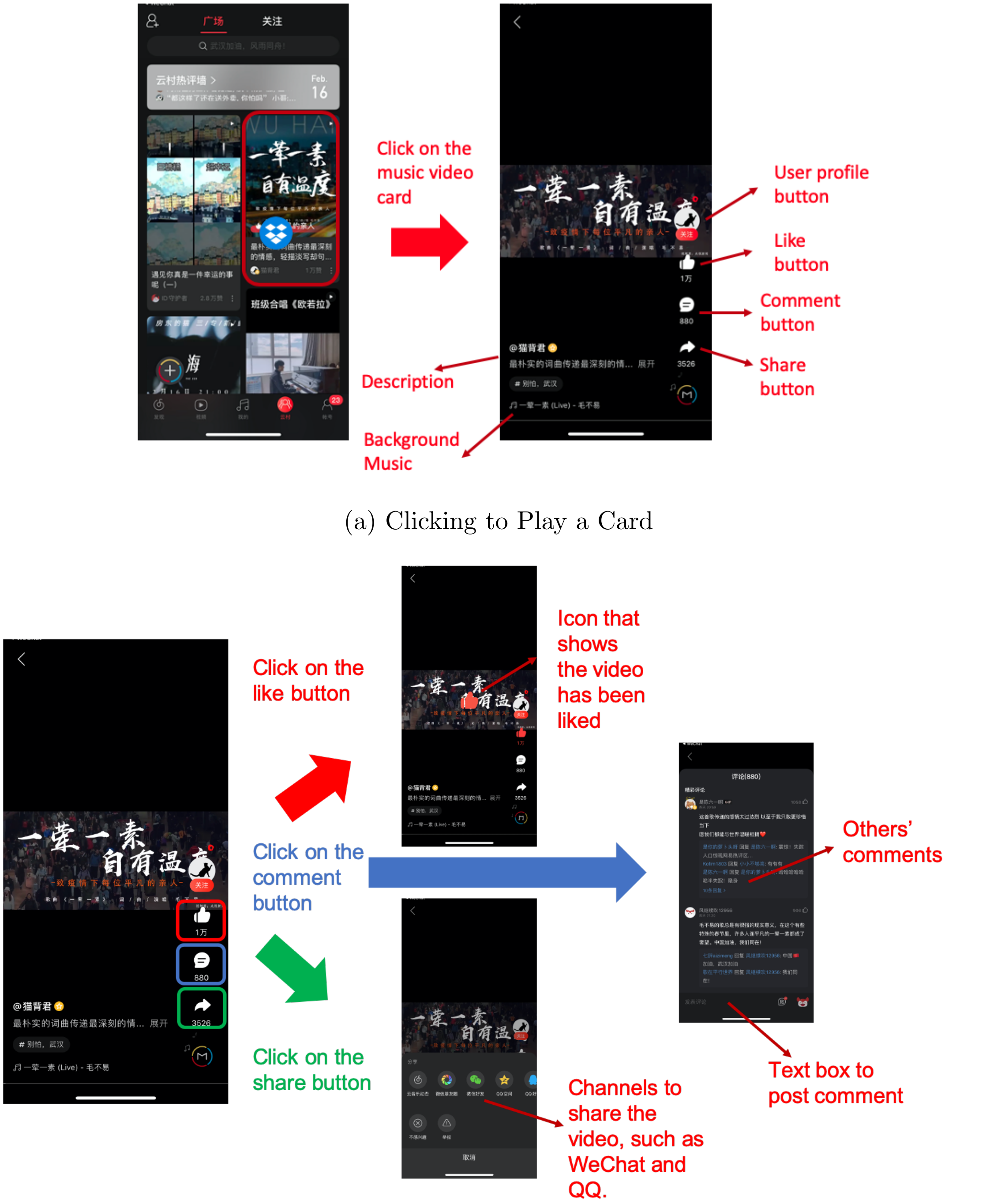
Table 2 Data Dictionary for impression_data.csv

Field Name	Data Type	Description	Sample Value
userId	string	The unique identifier of each user in the data set	MCPCHCMCHCIC
dt	numeric	The number of days from the start of the sample period	11
mlogId	string	The unique identifier of each card in the data set.	NCPCKCKCMCPCNC
impressTime	numeric	The epoch time of the impression	1574478123000
impressPosition	numeric	The position of impression in the feed	10
isClick	binary	1 if the user clicks on the card, 0 otherwise	1
isComment	binary	1 if the user comments on the card, 0 otherwise	0
isLike	binary	1 if the user likes the card, 0 otherwise	1
isShare	bianrdy	1 if the user shares the card, 0 otherwise	1
isViewComment	binary	1 if the user views comments from other users on the card, 0 otherwise	0
isIntoPersonalHomepage	binary	1 if the user enters the creator’s homepage through the card, 0 otherwise	0
mlogViewTime	numeric	The number of seconds that the user has spent on the card	136.09
detailMlogInfoList	string	JSON file contains all cards that the user sees if s/he swipes down	[{‘isZan’: ‘0’, ‘isComment’: ‘0’...][4]

is first shown to the user on her app (instead of the time when the user clicks on the impression). Each user may have multiple impressions in a given day; each card may be shown to multiple users during the sample period. Therefore, each row of impression data is uniquely identified by a combination of *userId*, *mlogId* and *impressTime*, representing that a card is shown to a user at a particular time. Each impression for a user comes with a position in her feed stream, and it is recorded in the *impressPosition* data field. The position starts with 1 and is counted from top to down and from left to right. In other words, if we have 4 cards on the app screen, as shown in Figure 2, the upper left has position 1, the upper right has position 2, the lower left has position 3, and the lower right has position 4.

For each impression, the table provides the users’ actions on the recommended cards. First, as shown in Figure 3a, a user could click on a card once the impression of the card is presented to the user. Once the user clicks on the card, the music video in the card will be automatically played in the user’s app in full screen mode if the card contains a video. If a card contains a set of images, the first image will be shown to the user in full screen mode. Such an action is recorded in the *isClick* data field. Second, Figure 3a also shows that, once clicking on a card, the user can comment and view other comments on the card. As shown in the middle scenario of Figure 3b, the users can view or post comments on a card by clicking on the comment section while watching the card. Once a user clicks on the comment tab, the comment section showing other people’s comments will appear along with a text box at the bottom. The user can view others’ comments or type in

Figure 3 Actions on a Card


广场
关注
Click on the music video card
User profile button
Like button
Comment button
Share button
Description
Background Music
1万
880
3526
@猫背君
Click on the like button
Icon that shows the video has been liked
Click on the comment button
Others' comments
Text box to post comment
Click on the share button
Channels to share the video, such as WeChat and QQ.

(a) Clicking to Play a Card

(b) Like, Comment, and Share a Card

textual information in the text box to post a comment. Moreover, Figure 3a displays that the total number of comments of a card is shown below the comment button to all users who have clicked on the card. Whether a user comments on a card for an impression is recorded in the *isComment* data field, and whether a user views others' comments on a recommended card is recorded in the *isViewComment* data field.

Third, a user can also like a card by clicking on the thumb-up button. In the upper scenario of Figure 3b, we can observe that, once the user clicks on the thumb-up button, the button will turn from white to red, and a big thumb-up logo will appear in the center of the screen for a couple of seconds to indicate that the user has successfully liked the card. Comparing Figure 2 and Figure 3a, we can see that the total number of likes of a card, unlike its total number of comments, is visible to all users, regardless of whether they have clicked on the card or not. Whether a user likes a card or not is recorded in the *isLike* data field. Fourth, a user can also click on the share button, as shown in the lower scenario of Figure 3b. Once a user clicks on the share button, a share screen will pop up from the bottom, asking the user which channel to share with. The user can share through various social media channels, such as WeChat and QQ. Whether a user decides to share the card from an impression, regardless of the channel, is recorded in the *isShare* data field. Moreover, a user can click on the creator's personal profile logo on top of the like button and get into the creator's personal page, and whether a user enters the creator's personal homepage from a card is recorded in the *isIntoPersonalHomepage* data field.

Last but not least, the users can also swipe down a card. Once a user decides to swipe down the card, a new card will be automatically recommended to the user and automatically shown in full screen mode. Notice that, if the current card has impression position n, the next card recommended to the user after swiping down may not have position $n+1$ since the user's actions of playing the current card would give the algorithm more information and update the algorithm's recommendation. If a user chooses to swipe down after watching a card recommended to her, the information regarding each of the cards shown after swiping down will be stored in the *detailMlogInfoList* data field of the focal card's data point in Table 2. The number of data points in *detailMlogInfoList* represents how many times that the user has swiped down after clicking on a card. Notice that the impressions through swiping down are different from those in the discovery subtab. A card will be automatically played if it comes from swiping down. But a card will only be played after being clicked if it comes from the discovery subtab. In summary, this impression-level data contains two sets of impression: (a) impression through the two streams of video cards in the tab, which is stored in each data point of Table 2, and (b) impression through swiping down, which is stored in the *detailMlogInfoList* data field in Table 2.

Furthermore, the table also provides the number of seconds for which the user has played the card. In other words, this is the difference between the time when the user clicks on the card and the time when the user swipes down or leaves the card by clicking the back button or closes the app. If the card contains a video and the user is still on the video page when the video ends, the

Table 3 Data Dictionary for mlog_demographics.csv and mlog_stats.csv

Panel A: Data Dictionary for mlog_demographics.csv

Field Name	Data Type	Description	Sample Value
mlogId	string	The unique identifier of each card in the data set.	NCPCKCKCMCPCNC
songId	string	The unique identifier of each song in the data set.	LCLCNCGCPCLCPCJCGC
artistId	string	The unique identifier of each artist of a song in the data set.	PCNCHCNCNCPCPCJC
creatorId	string	The unique identifier of each creator of a card in the data set.	KCJCKCNCNCLCLCICNC
publishTime	numeric	The number of days when the card is published till December 1st, 2019	109
type	binary	1 if the card contains a set of images and text with background music, 2 if the card contains a music video	1
contentId	numeric	The anonymized type of a card's content with 122 unique levels	500150125068
talkId	numeric	The anonymized topic of a card with $9,914$ unique levels	27004

Panel B: Data Dictionary for mlog_stats.csv

Field Name	Data Type	Description	Sample Value
mlogId	string	The unique identifier of each card in the data set.	NCPCKCKCMCPCNC
dt	numeric	The number of days from the start of the sample period	11
userImprssionCount	numeric	The number of unique users the card was shown to for a given date	133
userClickCount	numeric	The number of unique users who clicked on the card for a given date	65
userLikeCount	numeric	The number of unique users who liked on the card for a given date	8
userCommentCount	numeric	The number of unique users who commented on the card for a given date	1
userViewCommentCount	numeric	The number of unique users who viewed others' comments on the card for a given date	1
userShareCount	numeric	The number of unique users who shared the card for a given date	0
userIntoPersonalHomepageCount	numeric	The number of unique users who entered the creator's homepage from this card for a given date	0
userFollowCreatorCount	numeric	The number of unique users who followed the creator from the card for a given date	0

video will automatically replay. The watch time of an impression is recorded in the *mlogViewTime* data field. Notice that a user's total app usage cannot be imputed from this watch time since a user may browse other tabs in the app on a given day; and therefore, unfortunately, researchers cannot impute a user's total app usage in a day through this data set.

2.2. Card Data

We then introduce two data tables regarding each card in the impression data. Panel (A) of Table 3 describes the card demographics table which is at the card-level. This table offers information about each card that does not change over time. Each row in the table is uniquely identified by the *mlogId* data field, representing each card. For each card, we use the *songId* data field to identify the song used in the background. Each card can only have one song associated with it, but each song can be used for multiple cards. In fact, the most popular song in this data set has been used for 92,426 cards. For each card, we also provide the unique artist of the song used in the card, which is recorded in the *artistId* data field. The *creatorId* data field stores a unique identifier for each creator in the data set, and it can be used to join this card-level table with creator tables in Section 2.3. The *publishTime* data field represents the difference between the time when the card is initially published and the end of the sample period, which is December 1st, 2019. Panel (C) of Figure 4 shows the histogram of the *publishTime* data field. It can be seen that, even though the sample period is one-month-long, the sample contains cards that have been created before the sample period (i.e., cards with *publishTime* larger than 30). The *type* data field is important and differentiates a music video card from an image card. The *type* is 1 if the card contains a set of images and text with background music, and 2 if the card contains a music video. The *contentId* and *talkId* data fields represent anonymized categorical data related to each card. The *contentId* data field contains 122 levels representing the content category of the card, such as gaming or concert. The *talkId* data field has 9,914 levels indicating the hashtags used in the card, which is often about a particular event.

Besides this card-level table, the data set also includes a card-day-level table that provides the daily summary statistics of a card in the sample period. Panel (B) of Table 3 describes this data table. Notice that the summary statistics of a card does not only include the actions generated by users in our random sample (i.e., users in Table 5), but also include actions of all other users on the platform who have used the discovery subtab on the given date and interacted with the card. Each row in this table is uniquely identified by the combination of the *mlogId* data field and the *dt* data field, representing the summary statistics of a card for a given date.

We provide seven different summary statistics with respect to each card in a given day. The *userImprssionCount* refers to the total number of unique users a card was shown to in a given day. The *userClickCount* summarizes the total number of unique users who have clicked on a card during a given date. The *userLikeCount*, *userShareCount*, and *userCommentCount* data fields represent the number of unique users who have liked, shared, or commented on a card in a given day. Similarly, the *userViewCommentCount* and *userIntoPersonalHomepageCount* show the number of unique users who have browsed comments on or entered the creator's home page from a given card in a given day. Last, the *userFollowCreatorCount* shows the total number of unique users who have followed the creator of a card through this card in a day. Note that this number does not represent the total number of new followers that a creator generates in a day since users can also follow the creator from other cards and/or from other tabs.

Table 4 Data Dictionary for creator_demographics.csv and creator_stats.csv

Panel A: Data Dictionary for creator_demographics.csv

Field Name	Data Type	Description	Sample Value
creatorId	string	The unique identifier of each creator of a card in the data set.	KCJCKCNCNCLCLCICNC
gender	string	The predicted gender of the creator, which can be unknown or NA.	male
registeredMonthCnt	numeric	The number of months between this creator's registration time and December 1st, 2019	66
follows	numeric	The number of people a creator has followed on November 1st, 2019	66
followeds	numeric	The number of followers a creator has on November 1st, 2019	1
creatorType	numeric	The anonymized type of a creator with 10 levels	0
level	numeric	The activity intensity level of a creator ranging from 0 to 10	10

Panel B: Data Dictionary for creator_stats.csv

Field Name	Data Type	Description	Sample Value
creatorId	string	The unique identifier of each creator of a card in the data set.	KCJCKCNCNCLCLCICNC
dt	numeric	The number of days from the start of the sample period	11
PushlishMlogCnt	numeric	The number of cards this creator has created for a given date	1

2.3. Creator Data

Next, we discuss the two tables associated with each creator in the data set. The first table provides demographic information about each creator, and it is at the creator-level. The detailed data fields of this creator-level table are described in Panel (A) of Table 4. The unique identifier of this table is the *creatorId* data field. For each creator, we provide several key demographic information that could be revealed. First, we offer each creator's predicted gender in the *gender* data field, which can be either "male" or "female." This data field can also be either "NA" or "unknown," both of which represent that the gender of the creator is not known to the platform. Second, we provide the tenure of each creator (i.e., the number of months for which this creator has registered till December 1st, 2019), which is recorded in the *registeredMonthCnt* data field. Panel (B) of Figure 4 shows the histogram of this *registeredMonthCnt* data field, which indicates that the data set covers a wide set of creators with heterogeneous registration time on the platform. Third, we also provide each creator's number of followers and number of people she has followed by November 1st, 2019. This information is stored at the *followeds* and *follows* data fields. Fourth, the *creatorType* data field gives the anonymized genre of creators, which has 10 levels. Last but not least, the *level* data field represents the activity intensity of the creator ranging from 0 to 10. The activity intensity level is a combination of a user's app time and frequency of interactions with app; the smaller this number is, the less active the user is.

Table 5 Data Dictionary for user_demographics.csv

Field Name	Data Type	Description	Sample Value
userId	string	The unique identifier of each user in the data set	MCPCHCMCHCIC
province	string	The province in Pinyin that this user comes from	an hui
age	numeric	The predicted age of the user	21
gender	string	The predicted gender of the user	male
registeredMonthCnt	numeric	The number of months between a user's registration time and December 1st, 2019	66
followCnt	numeric	The number of people a user has followed till December 1st, 2019	1
level	numeric	The activity intensity level of a user ranging from 0 to 10	10

The second table in this category provides daily information about each creator, and it is described in Panel (B) of Table 4. Each row in this table is uniquely identified by the combination of the *creatorId* and *dt* data fields, which represents a unique creator in a given day. For each creator in each day, we provide the number of cards that this creator has created in that day. This information is recorded in the *PublishMlogCnt* data field.

2.4. User Data

The last part of our data consists of user demographic information for all users who appeared in our impression data set. Table 5 describes each data field of this user-level table. Each row in the table is uniquely identified by the *userId* data field. Similar to the creator-level table, we provide six key pieces of information regarding each user in our sample. First, we provide the province where the user resides in, which is recorded in the *province* data field. This data field can be any province in China or "NA," representing that the province of the user is not known. Second, the *age* and *gender* data fields provide the predicted age and gender of the user in our data set. Third, similar to the creator-level table, the *registeredMonthCnt* data field counts the number of months between this user's registration time and December 1st, 2019. Panel (A) of Figure 4 shows the histogram of the *registeredMonthCnt* data field. It can be seen that, even though the data sample is only one-month-long, it covers users with a wide spectrum of tenure on the platform. Fourth, we provide the number of people a user has followed till December 1st, 2019, in the *followCnt* data field. Last, we also provide the activity intensity level of each user in the *level* data field. Notice that, since a user can also post video on NetEase Cloud Village, a user can also be a creator. One can use userId in Table 5 to match with creatorId in Table 4 to get data of a user's creating and consumption activities at the same time.[5]

[5] Note that, since we only have 5% of users, not all creators in Table 4 in our data can be matched with a user. Similarly, not all users in Table 5 can be matched with a creator.

3. Conclusion and Data Access

This data set provided by NCM consists of six tables describing $2,085,533$ users' impression-level activities from the discovery subtab on the NCM app from November 1st, 2019 to November 30th, 2019. NCM and the RMP Section of INFORMS collectively invite researchers to conduct novel data-driven research in the field of revenue management and innovative marketplace analysis by using this data set. Among all possible research questions, the team at NCM who provides the data is mostly interested in the ones described in Section 1.

As mentioned above, these six tables in the data set can be divided into four categories. This first category is about the impression data and contains the impression table (i.e., impression_data.csv). This table records all 57 million impressions of cards that those $2,085,533$ users have experienced in the discovery subtab during the 30-day-long sample period. The second category is about cards and consists of the card-level table and the card-day-level table (i.e., mlog_demographics.csv and mlog_stats.csv). This information contains time-homogeneous and time-inhomogeneous daily information regrading each card in the impression data. The third category is about creators and contains two tables: one is at the creator-level (i.e., creator_demographics.csv), and the other is at the creator-day-level (i.e., creator_stats.csv). These tables contain information regarding each of the $90,534$ creators who appeared in our impression data. The last category has a user-level table, which provides user demographic information regarding each of the $2,085,533$ users in our impression data set.

We believe that this data set is unique in several ways compared to other public data sets in the literature and provides unique research opportunities. First, unlike previous data sets used in the recommendation literature, such as the Expedia dataset[6], this data set does not only contain what products had been recommended (i.e., the impression) but also the sequence of products that were recommended (i.e., the impression position). Such features allow the researchers to explore dynamic recommendation policies. Second, while the past literature on user-generated content typically focuses on user-day-level data (see, e.g., Zhang and Zhu 2011, Huang et al. 2019), this data set provides impression-level details. Such details could allows researchers to study how different micro-level dynamics affect the demand and production of user-generated content on a platform. Third, compared to traditional data sets used in two-sided platforms (see, e.g., Shen et al. 2019), this data covers not only the demand-side consumption but also the production-side response to demand feedback, such as likes, shares, and comments. Researchers can utilize this data to study typical two-sided control problems on a platform, such as matching, which otherwise can be difficult to study.

[6] https://www.kaggle.com/c/expedia-hotel-recommendations/overview

To access the data and participate in the competition, members of the RMP Section at INFORMS can go to the data hosting website on the INFORMS RMP Section website and follow the download link. The data file is a compressed file that contains all six tables in the CSV format.

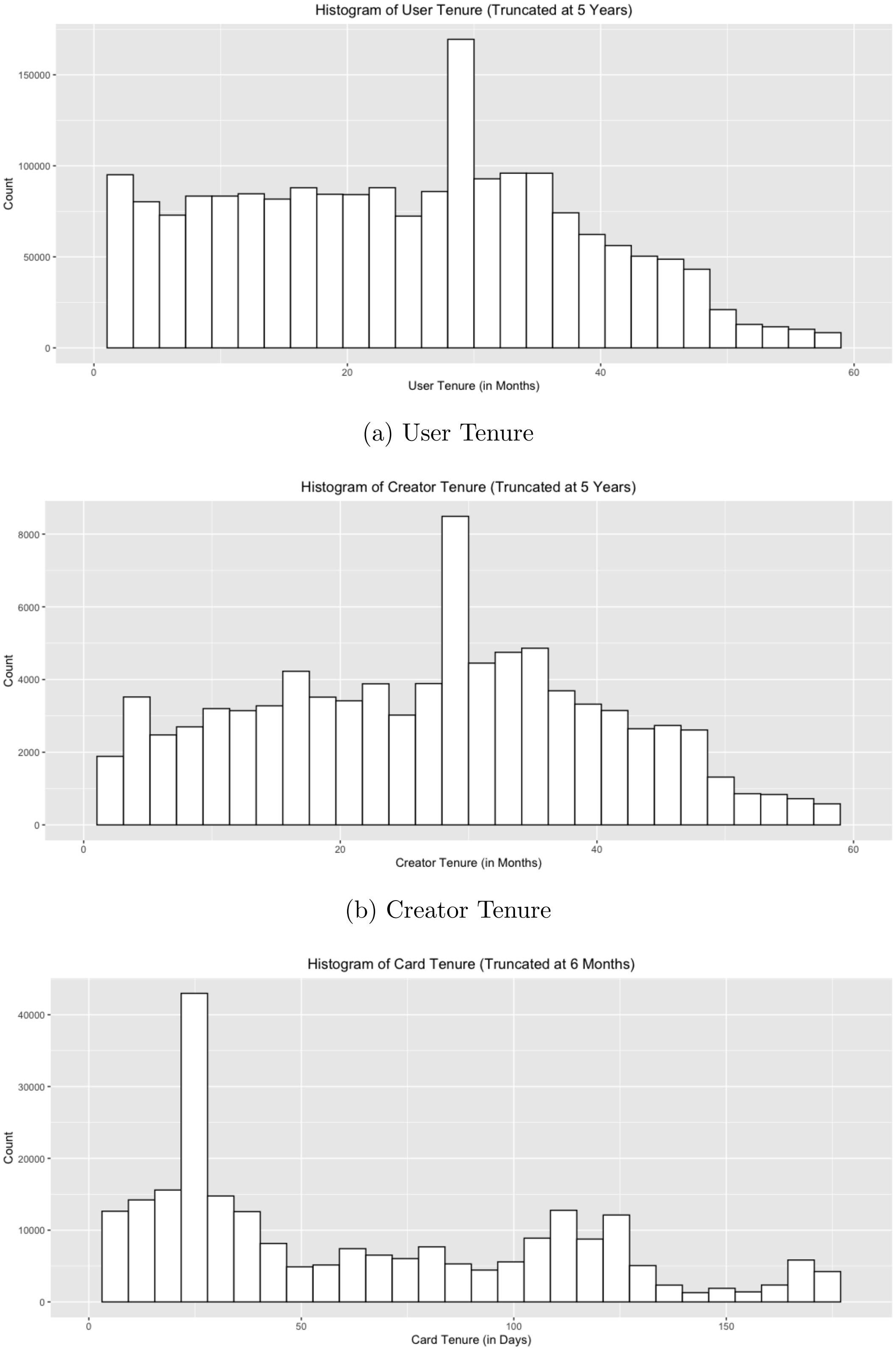
References

Huang, Ni, Gordon Burtch, Bin Gu, Yili Hong, Chen Liang, Kanliang Wang, Dongpu Fu, Bo Yang. 2019. Motivating user-generated content with performance feedback: Evidence from randomized field experiments. *Management Science* **65**(1) 327–345.

Shen, Zuo-Jun Max, Christopher S Tang, Di Wu, Rong Yuan, Wei Zhou. 2019. JD.com: Transaction level data for the 2020 MSOM data driven research challenge. *Available at SSRN 3511861* .

Zhang, Xiaoquan Michael, Feng Zhu. 2011. Group size and incentives to contribute: A natural experiment at Chinese Wikipedia. *American Economic Review* **101**(4) 1601–15.

Figure 4 Tenure of Users, Creators and Cards



(a) User Tenure

(b) Creator Tenure

(c) Card Tenure