

Case Study 4: Business Intelligence for Emerging Musicians

Group 1:

Janvi Kothari, Claire Danaher, Jonny Friedman, Renee Sweeney, Erin Teeple

December 7, 2017

DS 501

Introduction

Brand creation and growth present unique challenges in the age of the Internet. On the one hand, the Internet provides a multitude of opportunities for content distribution and access to global audiences. On the other, the speed, low cost of distribution, and finite attention span of any individual potential consumer of digital content present new, unique challenges for artists who seek to find an audience for their work through digital media. Specialized distribution platforms have been created to serve as points of exchange for various kinds of digital material. Examples of these types of sites include the movie rating site MovieLens³, which offers opportunities for movie viewers to exchange ratings and obtain personalized recommendations, and SoundCloud², which offers sound content creation, sharing, rating, curation, and recommendation functions.

For emerging artists without commercial support, however, there still exists the challenge of how to build their brand, manage their online present, and identify important target audiences and trends in a cost-effective way. Even though digital distribution and rating platforms can efficiently distribute content, these systems do not necessarily provide guidance or sufficient exposure for all new artists. Artists supported by business interests benefit from extensive support, including brand management teams comprised of image consultants, digital content advisors, and artistic mentors. In order to compete with more established artists supported by these resources, new artists need to find low-cost ways build their personal brand, manage their online presence, and stay on top of important cultural trends. Emerging artists could therefore benefit from a targeted data science service that reports information on regional music preferences and cultural trends related to their genre.

For this case study, we have developed a data science product that analyzes SoundCloud data to generate cost-effective business intelligence. The information available from our service can be used to identify distribution targets, refine branding, and enhance marketing of new music. We propose to create a website offering a country-wide interactive analysis summary of popular music genres by city location and summaries of genre-specific user-generated meta-data through analysis of SoundCloud tag data.

Business Plan

We present our business plan using the Business Model Canvas template.¹

Value proposition

We deliver insight to support strategic decision-making and to guide early musical career development and brand management. Information about how genre preferences differ by location is critical for identifying booking locations near as many potential fans as possible. The inclusion of user-generated meta-data in our data summaries provides further insight about genre-specific trends, popular cross-genre collaborations, and of-the-moment terminology. The service bundle we provide is access to a package of data visualizations.

Key Activities

Our key activities are SoundCloud user data analysis and distribution of our analysis summaries to subscribers using an online platform. Initially our business will be supported by site advertising revenues. As our users increase, we plan to change over to charging a subscription fee to access our site content.

Key Resources

We rely on publicly available SoundCloud user data to generate our analysis. Our distribution stream is through an online platform. Our initial revenue stream will be subscriber fees. The service has been designed with new artists as our target consumer.

Other potential consumers who might have an interest in our data might be agencies that represent early-career musical artists, restaurant or bar chain managers looking for guidance on musical styles to include in playlists for their establishments, and managers of performance venues trying to decide what talent to book.

Key Partners

Our current key partner is SoundCloud. We rely on SoundCloud as our data source. Future business development will include the identification of other data sources that could provide information on regional music preferences and consumption patterns.

Customer Relationships

Our customers will come to us for information to support their development as artists or for information on music consumer trends in their regions of interest.

Our customer segments expect us to establish a reliable, up-to-date information resource. This is the minimum viable product described in this case study. Additional cost-effective music brand development services might be helpful to offer through our web platform at some point in the future and these could be easily integrated.

Channels

Our customer segments want to be reached online. We offer a digital product. Advertisements online and at music festivals could be used to identify and reach potential customers.

Customer Segments

Our most important customers are emerging musical artists. We are creating value for emerging artists, agencies, and performance venues when we provide data-driven insights into regional trends in musical preferences and evolving terminology. We create value for our advertisers by attracting young artists for targeted advertising.

Cost Structure

Overhead costs will be relatively low. We require a website and use computing power to generate our updated analyses and mapping. SoundCloud data is currently freely available. The main cost is the expense of employing a data scientist/web content manager to maintain and update the site and to build on the current analysis product.

Revenue Streams

Revenue will be generated by advertising revenue from companies that offer goods and services targeted to new musical artists, for example recording and production businesses. After the site has an established audience, we plan to change over to charging a subscription fee for access.

Methods

SoundCloud is a social media platform for sound sharing². Users can interact with the site as sound “creators” who record or upload material; “listeners” who access audio from the site to listen to new music and receive track recommendations; and “curators” who listen and build playlists with the objective of gathering followers for their personalized recommendations. The SoundCloud API provides documentation and data. SoundCloud data can be searched and/or filtered using the SoundCloud API and keywords including tag_list, genre, title, and other information fields.

We downloaded data from the SoundCloud API using the q-parameter to search user profiles for the phrase “United States” to identify users living in the United States. We

then filtered this group of users specifically by country parameter = United States. This generated a list of 3000 unique individual users. Among these, we included in our final analysis only those users who had listed at least one favorite track. This yielded a total of 1500 unique users. For these 1500 users, we extracted the user-generated tags and genres of all favorite tracks and the user location by city. We then analyzed favorite genre and associated tags across multiple large cities in the United States (Table 1).

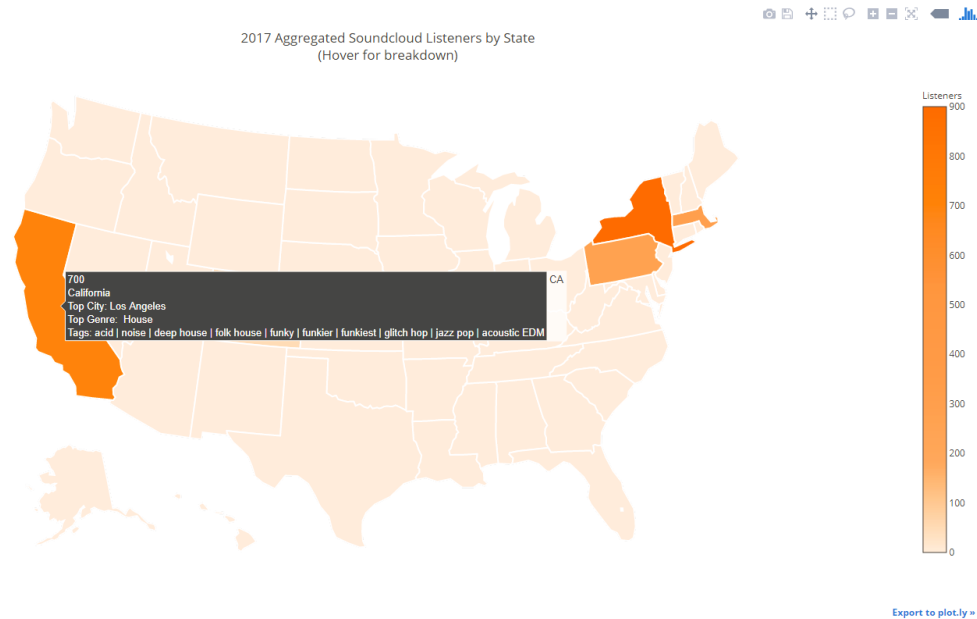
Table 1. Top genres and differentiating tags identified using TDF vectorizer class⁵.

City	Top Genre	Differentiating Tags
Chicago, IL	house	'house', 'techno', 'tech', 'remix', 'deep', 'music', 'edm', 'dj', 'electronic', 'bass'
Boston, MA	hip-hop & rap	'hop', 'rap', 'hip', 'hip hop', 'soul', 'hip hop rap', 'hop rap', 'dope', 'music', 'boston'
Miami, FL	trance	'trance', 'progressive', 'uplifting', 'uplifting trance', 'house', 'progressive house', 'pulsar', 'space', 'trance progressive', 'odyssey'
Houston, TX	hip-hop & rap	'rap', 'deen', 'remix', 'pop', 'bass', 'hop', 'soundcloud', 'music', 'hip', 'hip hop'
Nashville, TN	trap	'rock', 'hip', 'music', 'hip hop', 'hop', 'rap', 'bass', 'instrumental', 'trap', 'new'
Denver, CO	hip-hop & rap	'trap', 'rap', 'hop', 'hip', 'hip hop', 'music', 'rhymesick', 'bass', 'hiphop', 'fade'
Los Angeles, CA	house	'house', 'techno', 'pop', 'remix', 'tech', 'tech house', 'electronic', 'dance', 'trap', 'music'
New York, NY	house	'house', 'deep', 'deep house', 'music', 'remix', 'rap', 'hop', 'trap', 'electronic', 'hip'

For each user and location, we identify the genres most commonly marked as favorites and the tags associated with these favorite genres. On our website, we then report these by location on our site using an interactive map, along with the top tags associated with those genres. (Figure 1).

Tag analysis was performed using the TDF-IDF vectorizer. The TfidfVectorizer class provides a method for transforming a set of text documents into a TF-IDF feature matrix. TfidfVectorizer was used to construct of a classifier pipeline that filtered words (“tokens”) that were too common or infrequent in our text documents using TfidfVectorizer⁵. This method allowed us to ignore commonly used words while identifying other differentiating words occurring at high frequency in our samples tags.

Figure 1. Interactive map presenting most popular genres and top user-generated tags.



Analysis (the Math Part)

Problem formulations in Math: A chi-squared goodness of fit statistical test was run on the top genre in each city of our dataset. We wanted to test if our genre samples from SoundCloud were representative of the true music preferences within those cities. Since the true distributions are difficult to obtain, we assume that the genres are equally popular in each major city, but with more data from different sources, the expected distributions will become more accurate. The observed frequencies of the chi-squared test were how many times a city's top genre appeared and the city's total number of favorited tracks; the expected frequencies were 1 and the number of genres appearing in that city, converted to probabilities summing to 1. Another statistic we used was the Term frequency–inverse document frequency (TF-IDF), which calculates the importance of a term within a lexicon and its weight value. This helped us identify common and eye-catching language within the genre and location fields.

Math Solutions: As we expected due to the lack of data, our p-values for the chi-squared goodness of fit tests were close to 0, meaning that the observed and expected frequencies are not from the same distribution. In other words, our top genre is not a good representation of the actual top genre in each city. However, gathering better population data, possibly from web scraping other location-based music sales or streaming datasets, would yield better results in future work.

Zipf's Law was implemented to identify appropriate ranks for the top tags.

Implementation of the Solutions: Our team used the scipy library.

Discussion

We present here our business plan for an interactive analysis and visualization platform to inform new musician brand management. The information service we provide offers strategic insights to guide marketing, performance location scheduling, and trend engagement for early career sound artists. The problem of compiling and analyzing this type of information is important to solve because early career musicians have limited resources. Building a brand, attracting audiences, and maintaining a digital presence are significant challenges for modern sound artists. Established musicians are typically supported by multi-person brand management teams. Without the support of these resources, it is difficult to break out and achieve commercial success. The information we provide offers a competitive edge.

The difference between our data science approach and other music exchange and listening platforms currently available on the internet is that the focus of our site is on providing information that allows artists to identify regional genre preferences as the same time that we present user-generated tag meta-data. What tags are generated by music fans of the same genre in different cities? Are tags highlighting new cross-genre collaborations? Are tag themes similar across genres? These types of questions can be explored easily and intuitively using our visualization platform. Knowing this information, musicians can target audiences and define themselves in the context of the most important and current trends in the music industry. This idea certainly deserves investment. We provide business intelligence for emerging artists at a much lower cost than a conventional brand management team. Independent artists are important and unique voices. Each artist offers a unique perspective and creative vision. In order for each artist to have the best chance of success, they need insightful support. We propose to sell a product that provides that support as a competitive price point and with opportunities for build-out advertising and online sales services.

In the United States, the music industry generated \$3.9 billion dollars in 2016 alone, according to numbers reported by the Recording Industry Association of America (RIAA). Streaming services accounted for greater than 50% of this revenue. There exists an enormous market for digital music consumption – these practices both generate large volumes of consumer data for ongoing analysis and highlight the tremendous opportunities for artists who can create and grow a digital presence. We request from the sharks funding to support 6 months of salary for five full-time extremely talented data scientists and web site fees \$400,000.

REFERENCES

1. Business Model Canvas: nine business model building blocks, Osterwalder, Pigneur & al. 2010.
2. <https://soundcloud.com/>, accessed 12/7/2017.
3. <https://movielens.org/>, accessed 12/7/2017.
4. SoundCloud API <https://developers.soundcloud.com/docs/api/guide> , accessed 12/7/2017
5. http://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer.html, accessed 12/7/2017
6. <http://variety.com/2017/digital/news/streaming-services-us-music-revenue-2016-1202019504/>, accessed 12/7/2017