Characterizing Internet Radio Stations at Scale

Gustavo R. Lacerda Silva
Electrical Engineering Department
Universidade Federal de Minas Gerais
Belo Horizonte, Brazil
gustavolacerdas@ufmg.br

Lucas Machado de Oliveira
Computer Science Department
Universidade Federal de Minas Gerais
Belo Horizonte, Brazil
lucasmdo@dcc.ufmg.br

Rafael Ribeiro de Medeiros
Computer Science Department
Universidade Federal de Minas Gerais
Belo Horizonte, Brazil
medeirosribeiro@gmail.com

Olga Goussevskaia
Computer Science Department
Universidade Federal de Minas Gerais
Belo Horizonte, Brazil
olga@dcc.ufmg.br

Fabrício Benevenuto
Computer Science Department
Universidade Federal de Minas Gerais
Belo Horizonte, Brazil
fabricio@dcc.ufmg.br

ABSTRACT

In this paper we build and characterize a large-scale dataset of internet radio streams. More than 25 million snapshots of more than 75 thousand different radio stations were collected from the SHOUTcast service between December 2016 and April 2017. We characterized several attributes of the dataset, such as audience and music genre distributions among radio stations, advertisement and seasonal content dynamics, as well as bit rates and media formats of the radio streams. Finally, we analyzed to which extent these features affect audience size. We hope these and the other findings of our study may provide valuable information for content personalization and better advertisement placement in internet radio streams.

CCS CONCEPTS

• **Information systems** → *Multimedia information systems*;

KEYWORDS

Internet radio streams, radio popularity, Internet radio characterization, Multimedia Data Mining, Web Content Mining

ACM Reference format:
Gustavo R. Lacerda Silva, Lucas Machado de Oliveira, Rafael Ribeiro de Medeiros, Olga Goussevskaia, and Fabrício Benevenuto. 2017. Characterizing Internet Radio Stations at Scale. In *Proceedings of WI '17, Leipzig, Germany, August 23-26, 2017,* 8 pages.
https://doi.org/10.1145/3106426.3106540

1 INTRODUCTION

With the advent of the Internet, a new way of broadcasting audio content services has emerged [5, 20]. Also known as webcasting, radio transmission over the Internet made it possible not only for existing radio stations to transmit its content on a new platform but also helped the emergence of several independent radio stations [25].

One of the most popular free radio services is SHOUTcast[1]. During peak hours it has reached over 900,000 simultaneous listeners[2]. A key advantage of Internet radio over radio waves is the possibility of accessing a radio station from different countries, something that was not possible before due to the limited reach of electromagnetic radio waves. Another advantage is that a streaming Internet radio station is cheap to set up [25]. Moreover, if user-generated data is collected and analysed, more personalised content and better advertisement placement can be performed by the radio stations. In spite of the revolutionary transformation of Internet radio services, little is still known about Internet broadcasts. In fact, the availability of digitally-logged audiences about radio content provided by services like SHOUTcast is still very recent and basically unexplored. More importantly, it is still unclear what factors drive audience size in this kind of environment and to what extent radio stations display advertisements as part of their content streams. These are the key components that compose the financial model behind these services and understanding them is of interest to audio content producers, advertisement placement and, ultimately, of researchers exploring this research field.

In this work, we collected and characterized a large-scale dataset of Internet radio streams. To the best of our knowledge, this is the first work of this kind. We crawled the information provided by the SHOUTcast's Radio Directory API [2], collecting a total of over 25 million snapshots during time-windows between December 2016 and April 2017. Over 75 thousand different radio stations were detected in the collected dataset, of which 10% presented a mean daily audience of ≥ 10 listeners. Around $1,5\%$ of stations achieved a maximum daily audience of ≥ 100 listeners, and only 40 radio stations attracted ≥ 1000 listeners daily.

Firstly we analysed how music genres are distributed within the dataset. The majority of radio stations fall within a few popular music genres, such as Pop, Trip Hop, and Talk. The audience distribution among the music genres, however, is a little bit less uneven. We could detect that less common genres, such as Alternative and European, were at the top in terms of mean audience per radio station, suggesting that more specialized genres might attract more

[1] www.shoutcast.com
[2] www.webnethosting.net/kb/what-shoutcast-is

WI '17, August 23-26, 2017, Leipzig, Germany

ACM ISBN 978-1-4503-4951-2/17/08…$15.00
https://doi.org/10.1145/3106426.3106540

loyal listeners online, possibly because of the shortage of options of that kind of music on traditional radio stations.

A factor with influence on audience size that we detected was content seasonality. Since our collection time window included Christmas period, we could detect high audience peaks in radio stations with keywords related to Christmas in the station's name and metadata. We conjecture that seasonal content-oriented tagging can be used to boost audience. Moreover, the detection of seasonal influence in audience trends can provide valuable information for more personalized content and better advertisement placement in Internet radio stations.

Another attribute with impact on audience size was advertisement content dynamics. From the 75K collected radio stations, 61% contained at least one advertisement. We analyzed the prevalence of advertisement content in the collected radio streams and found out that radio stations with very high AD-to-NoAD content ratio had smal audience numbers.

Our characterization effort was mainly directed towards identifying factors that influence radio audience. Our analysis revealed some interesting findings. We ranked variable importance towards predicting audience size according to their Gini coefficient. Music genre and number of advertisements per hour turned out to be the most relevant features for predicting audience size on SHOUTcast.

The rest of this paper is organized as follows. Section 2 briefly surveys related work. Section 3 describes the methodology used in the data collection process. Section 4 discusses audience, genre, seasonal content, bit rate and media format-related characterization results. Section 5 analyzes advertisement prevalence in the radio streams. Section 6 presents the ranking of feature importance for prediction of audience size. Finally, in Section 7, we discuss our concluding remarks.

2 RELATED WORK

Research addressing Internet radio streaming has been conducted with different perspectives. Some studies focus on artist popularity analysis [4, 10]. Other works address user experience and quality of service [16, 18, 19]. Yet another exciting direction has been music recommendation [13, 24] and automatic playlist generation based on statistical analysis of radio streams [8, 9, 17, 24].

The emergence of various web-based music-related services has significantly changed the dynamics of the music industry. Faria et al. [10] realized that the traditional way to measure artist popularity, based on the number of disk sales and broadcasts on the radio, was not sufficient any longer. So they proposed a methodology for estimating artist popularity based on how they perform on different music digital media available on the Web, as well as data from traditional media, such as TV.

Bellogın et al. [4] also evaluated the popularity of artists considering web-based and social media music platforms. They studied the relationship between popularity indexes and their rankings. A dataset was used with 1,312 artists from three different social media music platforms (EchoNest, Last.fm and Spotify). The results indicated that popularity is more sensitive to the temporal dimension than service-dependent. The study also showed that the popularity is a rather stable signal for the majority of the indexes since it scarcely changes over time, and the ranking of popular artists is highly dependent on the index used to measure the popularity.

Melendi et al. [19] analyzed Internet radio services, considering the traffic between service devices, as well as different elements of user behaviour, such as resource consumption, quality of data transmissions, etc. In [18] Melendi et al. developed a simulation tool to perform evaluations of an Internet radio service, which was validated using a real service.

Turnbull et al. [24] explored the usage of personalized radio to promote the discovery of music by local artists. They studied `MegsRadio.fm`, a web-based service to stream a mix of music created by local and well-known artists based on seed artists, tags, venues and location. The results revealed that users become more open to listening to music created by local artists after using `MegsRadio.fm`.

Aizenberg et al. [1] used publicly available playlists of thousands of Internet radio stations to create a large-scale dataset and develop a probabilistic collaborative filtering model. Grant et al. [13] developed an Internet radio station recommendation system using historical data collected from the SHOUTcast platform.

Maillet et al [17] proposed an approach to generating steerable playlists from tags linked to songs collected from professional radio station streams. Chen et al [8] proposed to model playlists as Markov chains, generated through the Latent Markov Embedding (LME) machine learning algorithm, using online radio streams as a training set.

Finally, Küng et al. [15] categorized Internet radio into four different types: *(i)* the *Internet-stream* radio (which is the focus of this paper), in which the content is typically transmitted to the listener's device in real time; *(ii)* the time-shifted streamed radio, also known as *on-demand* audio; *(iii)* the time-shifted *downloaded* radio, in which the listeners download the content to their devices; and *(iv)* the *hybrid* service, which allows users to select a particular type of content.

To summarize, although there are many interesting efforts related to Internet Radio Services, none of them provides an in-depth characterization and understanding of audience size and advertisement dynamics in these systems. To the best of our knowledge, our effort is the first of this kind to tackle this issue.

3 METHODOLOGY AND DATA COLLECTION

In this section, we describe the source and the collection process of our dataset.

3.1 SHOUTcast Streaming Radio Service

SHOUTcast is a software for streaming media over the Internet. It was originally developed by Nullsoft and is now maintained by Radionomy [21]. SHOUTcast enabled the emergence of many internet radio stations, which are listed in a directory on its website [22].

SHOUTcast provides two different ways of broadcasting audio content: by streaming audio from its servers, or by downloading the software, installing and configuring it on a dedicated server. SHOUTcast keeps track of every station being transmitted and provides some services to them, such as detailed audience reports,

automated advertisement injection, geographically-sensitive content and monetization services.

SHOUTcast was chosen to be the subject of this study mostly because of the large number of radio stations being streamed continuously. According to the service's website, it has over 60,000 radio stations being broadcasted, which gives us a large amount of data to collect and analyze. Furthermore, it has a well-documented and easy-to-use API.

It also has a particular pattern to identify adverts, perhaps because of its monetization program, which injects adverts automatically into the radio streams. For its monetization service to work, advert audio files need to be configured in a specific way, which will be explained in greater detail in the next section.

3.2 Data collection

In this study, we collected the information provided by SHOUTcast's Radio Directory API. This API returns information about the current programming of a radio station, including radio station's name and ID, current track, audience size, media format and audio bit rate. We attached a timestamp to each record, to have a temporal overview of the data. To build our dataset, we developed a data crawler and hosted it on a cloud provider. The data was collected in 45-second intervals. Afterwards, the collected samples were aggregated in time intervals of one hour. The dataset used in the characterization process was thus comprised of a total of 25, 827, 411 records (75k radio stations), during the week of December 24^{th} to 31^{th} 2016 (8 days), and April 3 to 21^{th} 2017 (19 days).

3.3 Data limitations

Although the SHOUTcast API [2] provides an excellent opportunity to study online radio stations' activities, the collected dataset has some limitations.

Firstly, the API has a certain delay in updating the content being played. This was observed by listening to the audio stream content of 10 radio stations using the ff2mpeg software [11]. The observation took three days (12/26/2016 to 12/28/2016), and it was verified that, in some cases, even though the programming of the music or advert had changed, the API didn't reflect it instantly.

Secondly, the API doesn't share the geographic location of users and radio stations. This information is an important feature for analyzing advertisement placement [3] and discovering the right time zone of radio station programming.

Thirdly, some discontinuity in programming transmission was detected. To identify this behavior, the script was updated so it would generate a "record error" every time the radio station was unavailable for collection. Those errors resulted in a total of 2% of the collected records.

Finally, the API doesn't have a static unique ID to identify the radio stations. It has been observed that a radio station can disconnect at a given time and come back with a different identifier from what had been used previously. This impairs identification of each radio station, which was mitigated by using the radio station's name.

4 CHARACTERIZING AUDIENCE ON SHOUTCAST

To better understand how the audience of radio stations behaves and what influences it, we aim to analyze, in this section, the audience of SHOUTcast's radio stations regarding the following characteristics: mean audience per hour, radio station genres, seasonal content, media formats and bit rates.

4.1 Audience size

To have an initial understanding of the dataset, we first analyzed the overall audience of the radio stations. Figure 1 shows the CDF (Cumulative Distribution Function) of the mean audience per day, segmented into two groups: "All stations", representing all entries in the dataset, and "Top 100", representing the 100 most popular radio stations, measured by mean audience among the collected records. We also plot the maximum daily audience for the "All stations" group.

It can be seen that there is a small number of radio stations that have a high audience and a significant number of radio stations with a fairly small audience. In particular, approximately 90% of radio stations have a mean and maximum audience of ≤ 10 daily listeners. As SHOUTcast is a free service, anyone can create a radio station, which means that the vast majority of the stations have a low mean audience.

Nevertheless, 1% (approximately 750) and 1, 5% (1, 125) of the collected radio stations had a mean and maximum audience of ≥ 100 daily listeners, respectively. Moreover, if we zoom into the 100 most popular radio stations, we can see that more than 99% had an average of ≥ 100 listeners, and 40% had a mean daily audience of ≥ 1000 listeners.

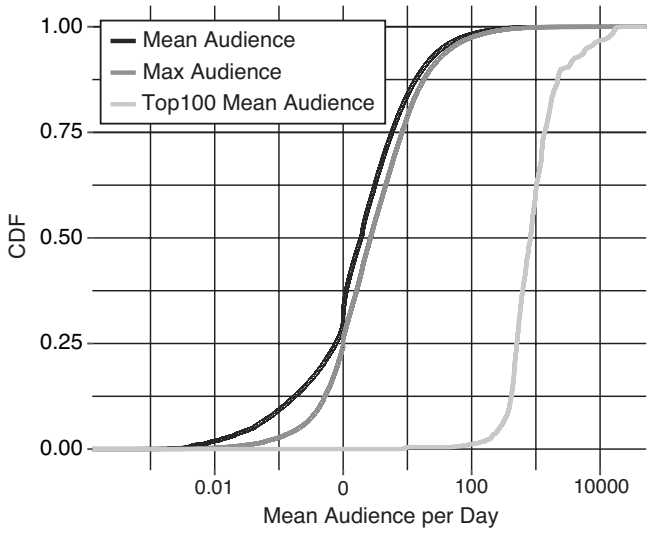

Figure 1: CDF of mean daily audience of a radio station on SHOUTcast.

In the following subsections, we analyze some characteristics that might affect the audience size of a radio station.

4.2 Music Genres

In this section, we focus on analyzing the distribution of music genres among the radio stations in our dataset. Figure 2 shows a list of the 15 most common genres in terms of the number of radio stations listing that genre in its metadata. Not surprisingly, Pop was the most frequent genre. Other popular genres can also be seen on the list, such as Talk, Rock, Gospel, Blues, and Dance.

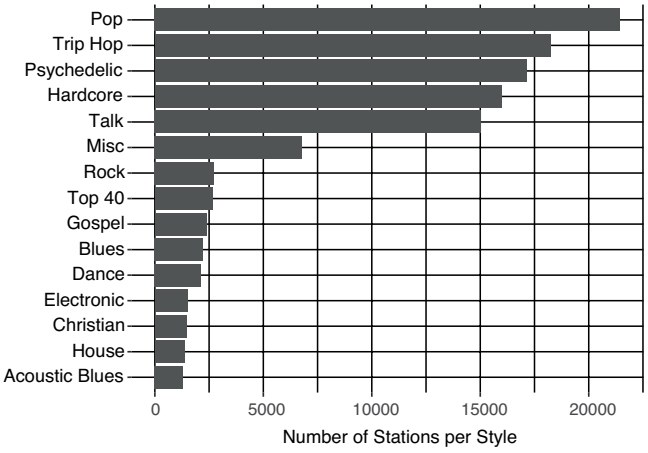

Figure 2: Top 15 genres by number of radio stations.

Figure 3 shows the list of the 15 most popular music genres in terms of mean audience per radio station. It can be seen that several genres from Figure 2 remain on the list. However, some popular genres drop to the bottom of the list, such as Pop and Rock, and others disappear altogether from the list, such as Gospel, Christian and Electronic. We can see that the Alternative genre not only makes into the list but is at the top of the list. This shows that some genres attract more listeners to specific radio stations, even though they are present in relatively fewer radio stations on SHOUTcast.

This suggests that, even though genres such as Pop and Rock are popular on SHOUTcast, the high competition, caused by numerous stations within that genre, in and outside SHOUTCAST, might cause many stations to have low and even zero audiences in that genre. On the other hand, less popular genres might attract more loyal listeners, possibly because of the shortage of options of that music style on traditional radio stations.

4.3 Seasonal Content

In December 2016 we detected some interesting dynamics in the audience of seasonal content. Figure 4 presents radio station names and their respective genre and audience during the first two days of the period analyzed (24th and 25th of December), while Figure 5 shows the same information, but considering the period of 26th to 31st of December. The increased audience of the "Merry Christmas" radio station was probably caused by the time of the year of our collection, which was around Christmas. Apart from "Merry Christmas" radio station, there are two other Christmas-themed stations in the top 15.

We can further detect the influence of seasonal content on the audience by comparing Figure 4 and Figure 5. We can see that

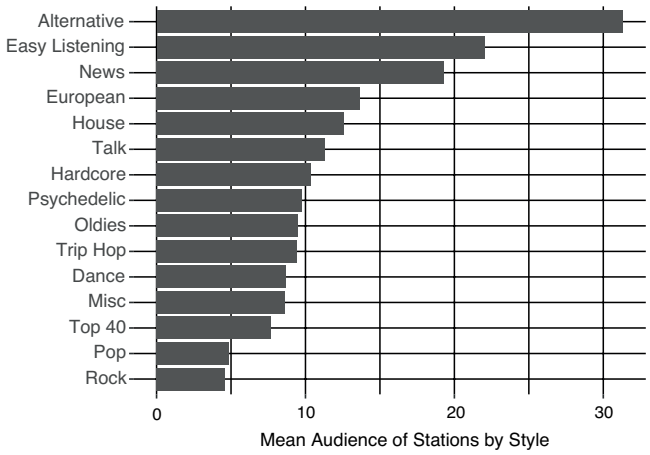

Figure 3: Top 15 genres by mean audience per radio station.

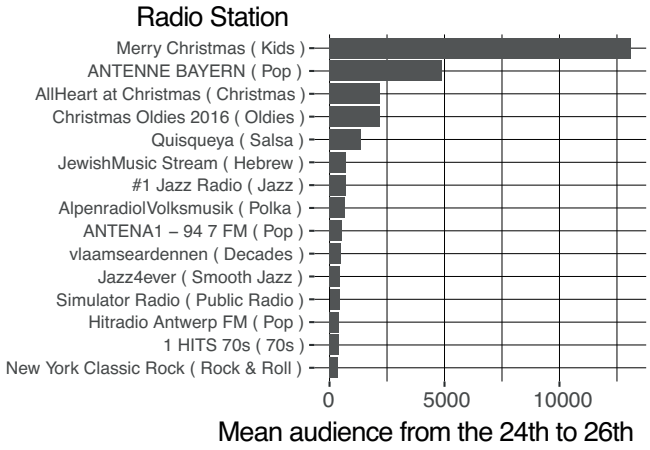

Figure 4: Top 15 radio stations and respective genres by audience from the 24th to 26th of December.

all of the Christmas stations that peaked on the 24th, 25th and 26th of December, no longer appear on the 15 most listened radios from 27th to 31st of December. During the former period, the most listened station had more the twice the audience of the second, and there were 3 Christmas-themed stations, while in the following period, all Christmas related stations disappeared from the top 15. This shows that certain seasonal events can significantly affect the audience of a radio station.

To corroborate our previous assumption, we analyzed the behavior of the audience of the three Christmas-themed stations that appear in the top 15. In Figure 6 we show the audience variation of all the Christmas-related radio stations among the top 15 during the collection period. It can be seen in all of them, that the audience peaked between the 24^{th} and 25^{th} of December and after that dropped significantly, indicating that seasonal content can, indeed, influence a station's audience.

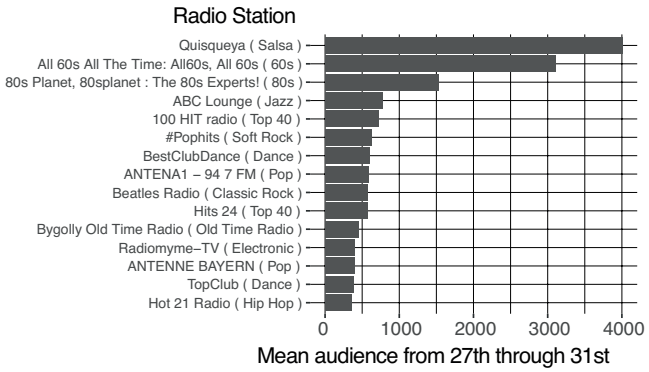

Figure 5: Top 15 radio stations and respective genres by audience from the 27th to 31st of December.

Based on previous analysis, we conducted a new experiment to verify if this behavior occurred in other music streaming platforms. Spotify Charts [23] is a website that compiles a list of the most streamed songs during a selected period on Spotify, which is one of the most popular music streaming services available nowadays. We analyzed the data considering the same period (12/24 to 12/31) and the results are presented in Table 1. It can be seen that Christmas-themed music streams have a peak in number during Christmas period, but right after that, their number drops significantly, maintaining a similar behavior to that detected in SHOUTcast's radio stations.

Table 1: Spotify streams during Christmas week.

Day	Total # streams	# Christmas songs	Percentage
12/24/16	215,904,792	58,758,925	27.21%
12/25/16	185,238,748	54,390,411	29.36%
12/26/16	147,661,069	74,808,54	5.06%
12/27/16	160,904,262	1,511,056	0.93%
12/28/16	167,354,900	956,865	0.57%
12/29/16	169,152,480	0	0%
12/30/16	173,237,247	0	0%
12/31/16	208,083,869	0	0%

Of all the Christmas-themed music streams on Spotify, the most played track was Mariah Carey's "All I Want For Christmas Is You", with a total of 11.429.780 streams during the analyzed period.

To further corroborate the previous behavior, we examined the song's number of views on YouTube, another popular content streaming service. As we can see in Figure 7, throughout most of the year, the video had a relatively small number of views, peaking on Christmas day and dropping significantly right after it. This behavior happens consistently in each of the three years covered in Figure 7.

As a matter of fact, seasonal content increased the audience of not only content-related radio stations, but the overall audience throughout the entire dataset was affected. As we can see in Figure 8, the overall audience peaked during Christmas and dropped significantly after that, rising again on New Year's Eve, which shows a tendency of users to listen to the radio during festive dates.

4.4 Media formats and bit rates

SHOUTcast allows radio stations to stream their content using different bit rates and two different media formats (MPEG and AACP). The bit rate influences the quality of the stream: the higher the bit rate, the higher the quality, which also implies a bigger volume of data to be streamed, requiring more bandwidth consumption.

Figures 9 and 10 show the distribution of different bit rates according to the number of radio stations and mean audience of the stations, respectively. We can see that, even though the vast majority of stations transmit at 128 bits per second, stations transmitting at 120 bits have a higher mean audience, which might indicate a user preference towards less bandwidth-consuming data streams.

The stronger adoption of slightly lower bit rates by users might be related to bandwidth consumption. Users might be interested in balancing quality and consumption, when listening to radio over slow Internet connections, such as mobile phones, so it is possible to have a decent audio quality without consuming too much data.

In the dataset, there is a considerable amount of stations transmitting at more than one bitrate. Even though they are treated as different stations by SHOUTcast's system, it is possible to detect this by checking the station's name. This decision is understandable if a station intends to reach a broader audience, which includes users streaming on mobile phones with limited Internet bandwidth, as well as users streaming from computers with broadband Internet connections.

Regarding the media format, SHOUTcast's service supports only 2 types of codecs: MP3 and AAC. According to Brandenburg [6], MP3 (short for MPEG-1 Layer-3) and ACC (short for MPEG-2 Advanced Audio Coding) are both audio encoders, with the latter being an advanced version of the former. Nevertheless, MP3, which was developed in 1991, is still more widespread than AAC on SHOUTcast.

Figure 11 presents the number of radio stations and mean audience by each media format, respectively. As we can see, the majority of radio stations choose to stream its content using MP3. Interestingly, when it comes to the average audience, the difference is not as high. This can be explained, again, by the fact that most stations using MP3 have small audience.

5 ADVERTISEMENTS ON SHOUTCAST

Advertisement placement in any media channel is a challenge (content, runtime, frequency, etc.) since bad ad placement can cause poor user experience. SHOUTcast's case isn't different. Advertisement is frequently a radio's primary source of income, so it is important for both stations and advertisers to optimize its ad content placement.

SHOUTcast provides users with an automatic ad injection system, which takes into account variables such as geographic location and time of the day to better position advertisement. For this system to work, ad audio files must be configured in a specific way: the track and artist name on the file's metadata must be set to "Advert:" [2]. This pattern makes it easy to identify this type of ad content in the dataset.

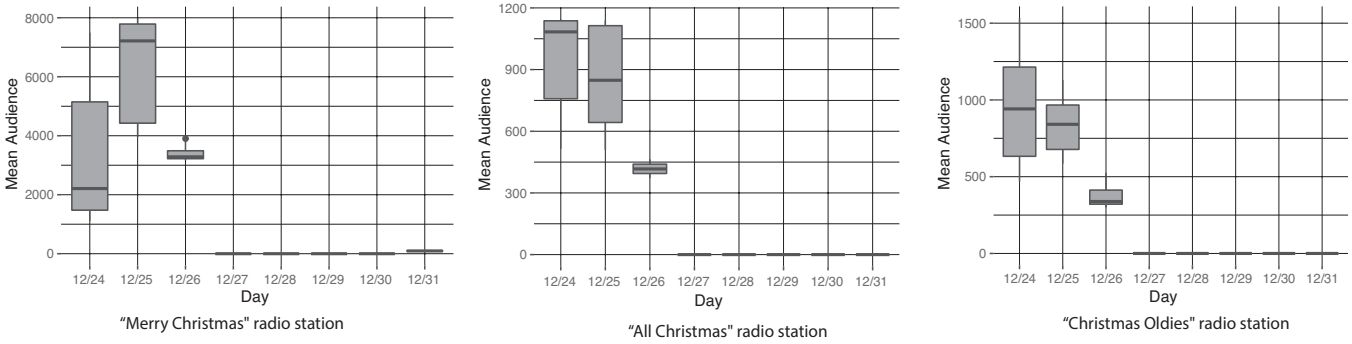

Figure 6: Audience dynamics of three radio stations before and after Christmas

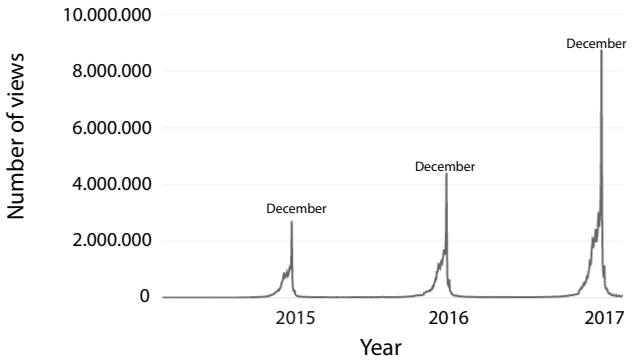

Figure 7: "Mariah Carey - All I Want For Christmas Is You" views on YouTube in the last 3 years.

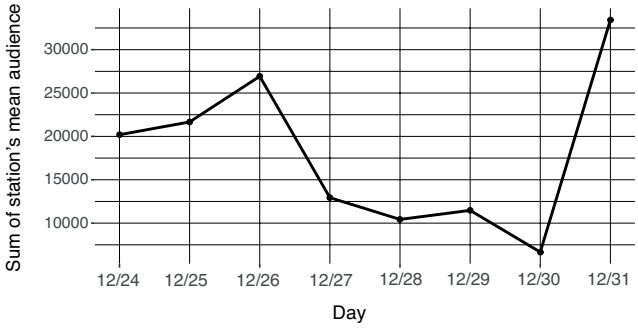

Figure 8: Overall dataset audience during the collected period.

To better understand how advertisement affects a station's audience, we divided the dataset into two groups: AD and NoAD. The first represents the advertisements, while the second represents every other content. A ratio > 1 means that the station has more advertisement content than non-advertisement content, while a ratio < 1 means the stations has more non-ad content. A ratio of exactly 1 means the amount of ad and non-ad content is the same.

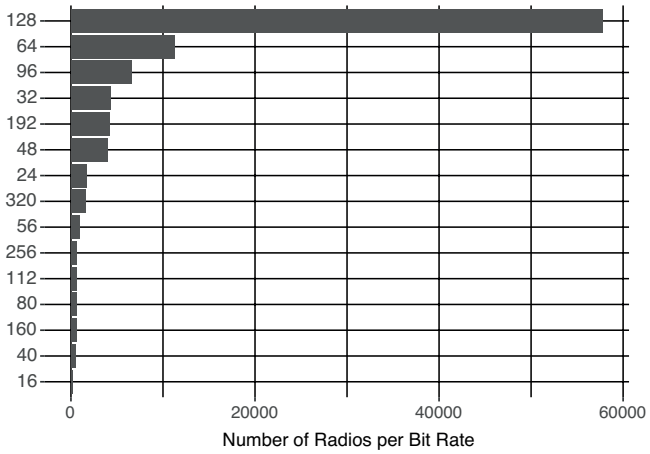

Figure 9: Bit rates popularity, by number of radio stations.

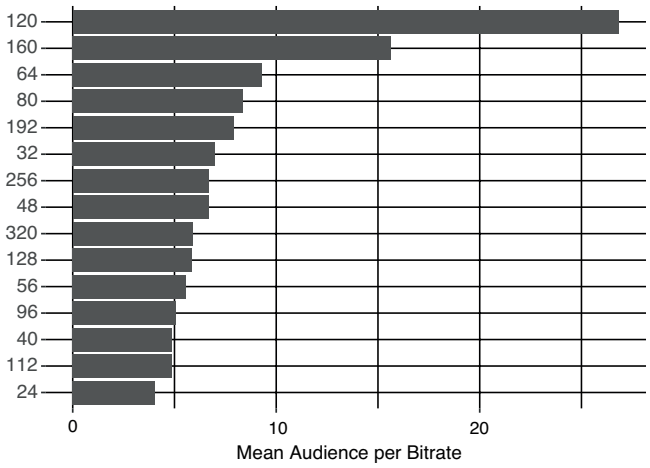

Figure 10: Bit rates popularity, by audience.

Figure 12 presents the CDF of the AD to NoAD ratio in our dataset. As we can see, 38.1% of all stations have zero advertisements in their schedules, and around 62.8% of the stations have a ratio

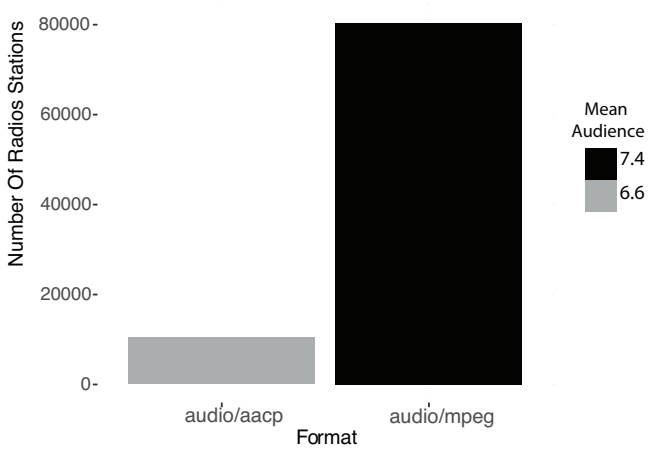

Figure 11: Number of radio stations and mean audience by media type.

≤ 0.01, which means that the vast majority of radio stations don't stream a lot or any ad content. Since SHOUTcast is a free and open source, there is a high number of independent radio stations, that may not be big enough for companies to be interested in advertising on them, which might explain the low occurrence of advertisement in the majority of the stations. Nevertheless, 3% of radio stations (approximately 2, 250) have a ratio ≥ 1, which means half of the content they stream is comprised of advertisement.

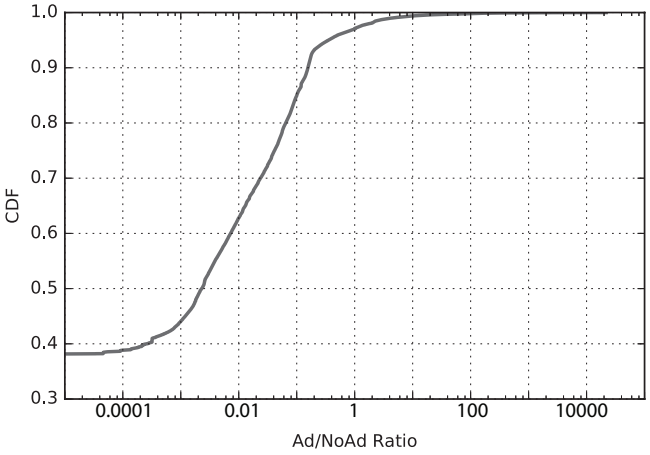

Figure 12: CDF of AD/NoAD ratio among radio stations.

Figure 13 illustrates how the AD/NoAD ratio correlates with the average audience of a radio station. As we can see, stations with a high average audience, have a low AD/NoAD ratio. In fact, the median ratio is 0.00229, which means, as said before, that most stations are streaming much less advertisement than other types of content.

To complement the analysis, Pearson's correlation between the AD/NoAD ratio and the audience was measured ($-1.298452e-05$). The correlation was detected with a 95% confidence interval. Despite the quite small number, the negative correlation indicates that the higher the Ad/NoAd relation, the lower the audience, which means that increasing the amount of ad might negatively affect the audience. In fact, we can see in Figure 13 that radio stations with very high Ad/NoAd ratio, have near zero audiences.

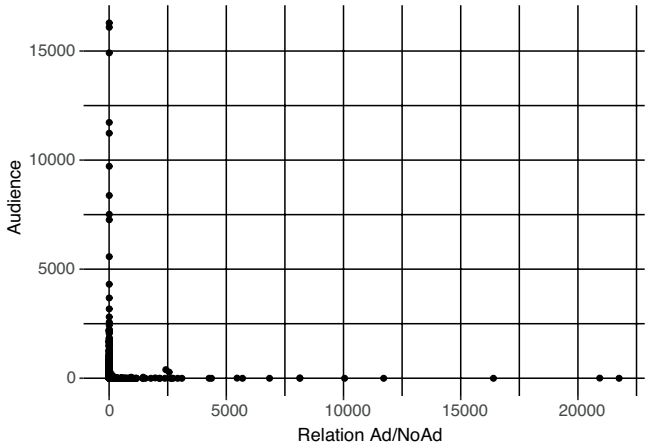

Figure 13: Ad/NoAd ratio by Mean Audience.

6 FEATURE RELEVANCE

In the previous sections, we analyzed the behavior of the stations' audience by a number of different features, such as genre, time of the year, media format and so on. Those analyses helped us to have a better overview of the dataset. To further understand what influences the audience of a station, we used our dataset to create a classifier, using machine learning techniques, to calculate the feature importance.

For this, we modeled our data as a classification problem, where the target is the mean audience and the data consists of the station's genre, the number of ads per hour, the number of records per hour, bit rate, media type, the hour of the day and day of the week. To train our classifier, we used Extremely Randomized Trees, as proposed by Geurts et al. [12]. This method is based on the Random Forest method, which operates by constructing a number of decision trees at training time and outputting the class that is the mode of the classes of the individual trees. The first algorithm for random forests was created by Tin Kam Ho [14] and was later extended by Breiman et al. [7].

This algorithm uses the Gini impurity index for the calculation of divisions during training. According to Breiman et al. [7], "every time a split of a node is made on variable M, the Gini impurity criterion for the two descendant nodes is less then the parent node. Adding up the Gini decreases for each individual variable over all trees in the forest, gives a fast variable importance that is often very consistent with the permutation importance measure".

Figure 14 shows the importance of each feature, calculated by the classification algorithm described above. The bars represent the feature importances of the forest, along with their inter-trees variability. The model has an accuracy of 0.89.

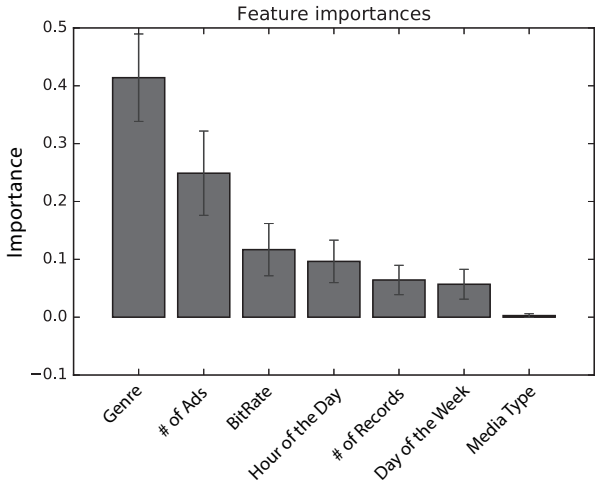

Figure 14: Features relevance

As we can see, the feature that influences the mean audience the most is the station's genre, with a Gini importance index of over 0.5, followed by the number of ads per hour. The media type is the least informative feature when it comes to audience.

This shows us that the music style that is played by a station can, indeed, have a high impact on the station's audience, with the number of ads being less but still influential.

7 CONCLUSIONS

In this paper we collected, analyzed, and presented a characterization of SHOUTcast's Radio Directory. To the extent of our knowledge, this is the first large-scale characterization effort of Internet radio streams, and, in particular, of the SHOUTcast service.

We collected and processed more than 25 million snapshots of more than 75 thousand different radio stations. As expected, even though most of the radio stations have low popularity, some achieve high audience numbers ($\geq$ 4000 listeners per hour). We characterized several attributes of the dataset, such as audience and music genre distributions among radio stations, advertisement and seasonal content dynamics, as well as bit rates and media formats of the radio streams.

Finally, we analyzed to which extent these features affect audience size. We ranked variable importance towards predicting audience size according to their Gini coefficient. Music genre and the number of advertisements per hour turned out to be the most relevant features towards predicting audience size on SHOUTcast, followed by the bit rate and time of day. As opposed to bit rate, media format presented no relevance towards audience prediction.

Moreover, we could detect that seasonal content can significantly affect a radio station's audience. We conjecture that, because of that, search words and tags can be used to strategically boost the number of listeners. The detection of seasonal influence in audience trends can provide valuable information for more personalized content and better advertisement placement in Internet radio stations.

Finally, we also intend to openly release the collected dataset for the community, at a later date.

ACKNOWLEDGMENTS

The authors would like to thank CNPq, CAPES and FAPEMIG for the financial support

REFERENCES

[1] Natalie Aizenberg, Yehuda Koren, and Oren Somekh. 2012. Build Your Own Music Recommender by Modeling Internet Radio Streams. In *Proceedings of the 21st International Conference on World Wide Web (WWW '12)*. ACM, New York, NY, USA, 1–10. https://doi.org/10.1145/2187836.2187838

[2] SHOUTcast API. 2017. SHOUTcast API. (Jan. 2017). https://www.shoutcast.com/Developer

[3] Shannon P Bauman, Keith Schmidt, and Dominic Preuss. 2006. Location Based, Content Targeted Online Advertising. (Oct. 5 2006). US Patent App. 11/539,109.

[4] Alejandro Bellogın, Arjen P de Vries, and Jiyin He. 2013. Artist popularity: do web and social music services agree. In *Int. Conf. on Weblogs and Social Media (ICWSM), Boston.*

[5] David Black. 2001. Internet radio: a case study in medium specificity. *MEDIA CULTURE AND SOCIETY* 23, 3 (2001), 397–408.

[6] Karlheinz Brandenburg. 1999. MP3 and AAC explained. In *Audio Engineering Society Conference: 17th International Conference: High-Quality Audio Coding*. Audio Engineering Society.

[7] Leo Breiman. 2001. Random Forests. *Machine Learning* 45, 1 (2001), 5–32. https://doi.org/10.1023/A:1010933404324

[8] Shuo Chen, Josh L Moore, Douglas Turnbull, and Thorsten Joachims. 2012. Playlist prediction via metric embedding. In *18th ACM SIGKDD.*

[9] Shuo Chen, Jiexun Xu, and Thorsten Joachims. 2013. Multi-space probabilistic sequence modeling. In *19th ACM SIGKDD.*

[10] Felipe Lopes de Melo Faria, Débora M. B. Paiva, and Álvaro R. Pereira. 2016. *ArtistRank – Analysis and Comparison of Artists Through the Characterization Data from Different Sources.* Springer International Publishing, Cham, 60–76. https://doi.org/10.1007/978-3-319-42092-9_6

[11] ffmpeg. 2017. ffmpeg. (Jan. 2017). https://ffmpeg.org/

[12] Pierre Geurts, Damien Ernst, and Louis Wehenkel. 2006. Extremely randomized trees. *Machine learning* 63, 1 (2006), 3–42.

[13] Maurice Grant, Adeesha Ekanayake, and Douglas Turnbull. 2013. MeUse: Recommending Internet Radio Stations.. In *ISMIR.* 281–286.

[14] Tin Kam Ho. 1995. Random decision forests. In *Document Analysis and Recognition, 1995., Proceedings of the Third International Conference on*, Vol. 1. IEEE, 278–282.

[15] Lucy Küng, Robert G Picard, and Ruth Towse. 2008. *The internet and the mass media.* Sage.

[16] Jin Ha Lee, Yea-Seul Kim, and Chris Hubbles. 2016. A look at the cloud from both sides now: An analysis of cloud music service usage. In *Proceedings of the 16th International Society for Music Information Retrieval Conference. New York: ISMIR.*

[17] François Maillet, Douglas Eck, Guillaume Desjardins, and Paul Lamere. 2009. Steerable Playlist Generation by Learning Song Similarity from Radio Station Playlists. In *ISMIR.*

[18] David Melendi, Roberto García, Xabiel G Pañeda, Sergio Cabrero, and Victor García. 2010. Performance Evaluation of Different Architectures for an Internet Radio Service Deployed on an Fttx Network. *International Journal of Business Data Communications and Networking (IJBDCN)* 6, 2 (2010), 46–68.

[19] D. Melendi, M. Vilas, R. Garcia, X. G. Paneda, and V. Garcia. 2006. Characterization of a Real Internet Radio Service. In *32nd EUROMICRO Conference on Software Engineering and Advanced Applications (EUROMICRO'06).* 356–363. https://doi.org/10.1109/EUROMICRO.2006.28

[20] Chris Priestman. 2002. *Web radio: radio production for internet streaming.* Gulf Professional Publishing.

[21] Radionomy. 2017. Radionomy. (Jan. 2017). https://www.radionomy.com/

[22] SHOUTcast. 2017. SHOUTcast. (Jan. 2017). https://www.shoutcast.com/

[23] Spotify. 2017. Spotify. (Jan. 2017). https://spotifycharts.com/

[24] Douglas R. Turnbull, Justin A. Zupnick, Kristofer B. Stensland, Andrew R. Horwitz, Alexander J. Wolf, Alexander E. Spirgel, Stephen P. Meyerhofer, and Thorsten Joachims. 2014. Using Personalized Radio to Enhance Local Music Discovery. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems (CHI EA '14).* ACM, New York, NY, USA, 2023–2028. https://doi.org/10.1145/2559206.2581246

[25] Tim Wall. 2004. The political economy of Internet music radio. *Radio Journal: International Studies in Broadcast & Audio Media* 2, 1 (2004), 27–44.