Table of Contents

ABSTRACT

With the visualizations in this project, the objective is to analyze how we as humans listen to music and how this information can be used for the growth of an up and coming artist in the music field. Big data analytics plays a huge role in determining the success of an entity. It allows for a thorough knowledge of who the audience is and how more people can be added to this audience. It also allows for a product or service to be improved by analyzing customer behaviour towards their product/service.

Following this, the music industry is a huge and saturated industry yet there have been, and still are, musicians that gain popularity overnight, and all this is possible majorly through big data analytics. In this project, I attempt to apply analytical techniques on public datasets which contain streaming services' users' listening activity to figure out ways in which music that they like can be delivered to them better.

Using a dataset of multiple users' listening activity, the aim is to obtain concluding information on demographic, time-related and other factors for listeners and subsequently analyze and formulate business plans for artists.

INTRODUCTION

Background:

Marketing and big data: two buzzwords that are used very frequently in today's world, and for good reason. The marketing department of a corporation is responsible for maximizing outreach with an appropriate audience to subsequently maximize growth and success of the organization. On the other hand, big data is defined as a large accumulation of information consisting of varied data types that can be used as and when needed. So, why are these two terms used together? Data can be in many forms, one of them being data collected from users of a particular product/service. This serves as a goldmine for marketing professionals as a lot can be said and predicted about a product/service just by customer behavior which is represented through the above-mentioned data. For example, marketers can identify a "target" audience, a population sample that is most likely to respond positively to their service, through their digital activity.

When it comes to music, there is a plethora of data that can be utilized for the publicity and benefit of musicians. Some examples are overnight hits such as "Habits" by "Tove Lo" and "Old Town Road" by "Lil Nas X". Tove Lo, managed by Ankit Desai at the time of this song's release, was primarily based in Sweden. Through his company Snaffu Records, that works on AI in music, Desai was able to see potential in Tove Lo being a success through numbers and a thorough analysis of Sweden's population and how they listen to music. Tove Lo later collaborated with top DJs and eventually climbed the ranks to #1 in Sweden. Similarly, even though Lil Nas X was banned from the Billboard and SoundCloud listings on account of incorrectly categorizing his song as “country”, regardless he gained high traction and went on to break multiple records solely through targeting people and locations that were most probable to take a liking to his country-trap music.

However, with great power comes great responsibility. According to the recent statistics by Spotify, they reported close to 40,000 new tracks being uploaded every day; and this is just one media platform. Clearly, the music industry is a saturated one and getting to the top requires strong marketing strategies. One strategy, which is often employed in almost every field, is to *know* the audience and deliver the product accordingly. Similarly, a musician would obviously want their music to be as widespread among the public as possible, but more importantly they would want to reach their intended audience first. For example, showing an ad for a new release by popular country artist Kenny Rogers to 18 year olds is probably not the best of ideas, however showing them the latest release by Post Malone would probably land someone in the marketing team a promotion. Choosing the right platform to market albums on also plays a huge role in determining the success of a musician. The most recent data and it’s corresponding visualizations figure prominently when tracking digital popularity of platforms among a widespread user base. Another thing that commands popularity of a track, as weird as it sounds, is the track length. This is more prevalent in today’s digital age where people have a very short attention span and if the song doesn’t seem to be interesting enough within seconds of their first listen, it’s not likely to be a major hit. A popular marketing strategy to tackle this is to use a portion of the song tailored specifically for a target audience and use it for an advertisement.

Motivation:

As discussed above, there are a plethora of ways in which music marketing can be revolutionized using big data and appropriate visualization and analysis. There are quite a few corporations and individuals who work actively in this field and the use of big data in music is a fairly new concept that is still being perfected every day through thorough research and application. However, there is still a lot of scope to improve upon this and make it more localized and accessible to smaller artists as they need maximum growth.

With that being said, this project aims to explore the different ways in which human psychology can be influenced by artist managers in the music industry to gain traction and popularity. Consumers want products in as simple of a manner as possible, and presenting this to them in the right way can be quite a task. However, there are quite a lot of interesting methods and techniques already implemented in the industry which this project goes through.

Pre-existing Visualizations:

For a start, here are a few visualizations that represent a typical user's listening activity

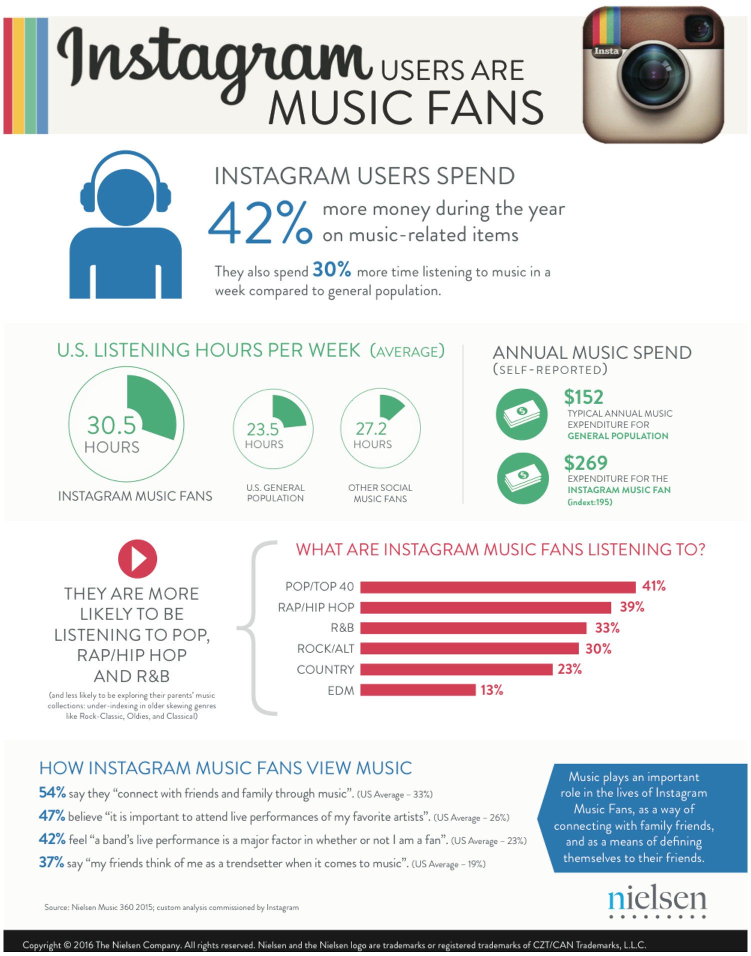
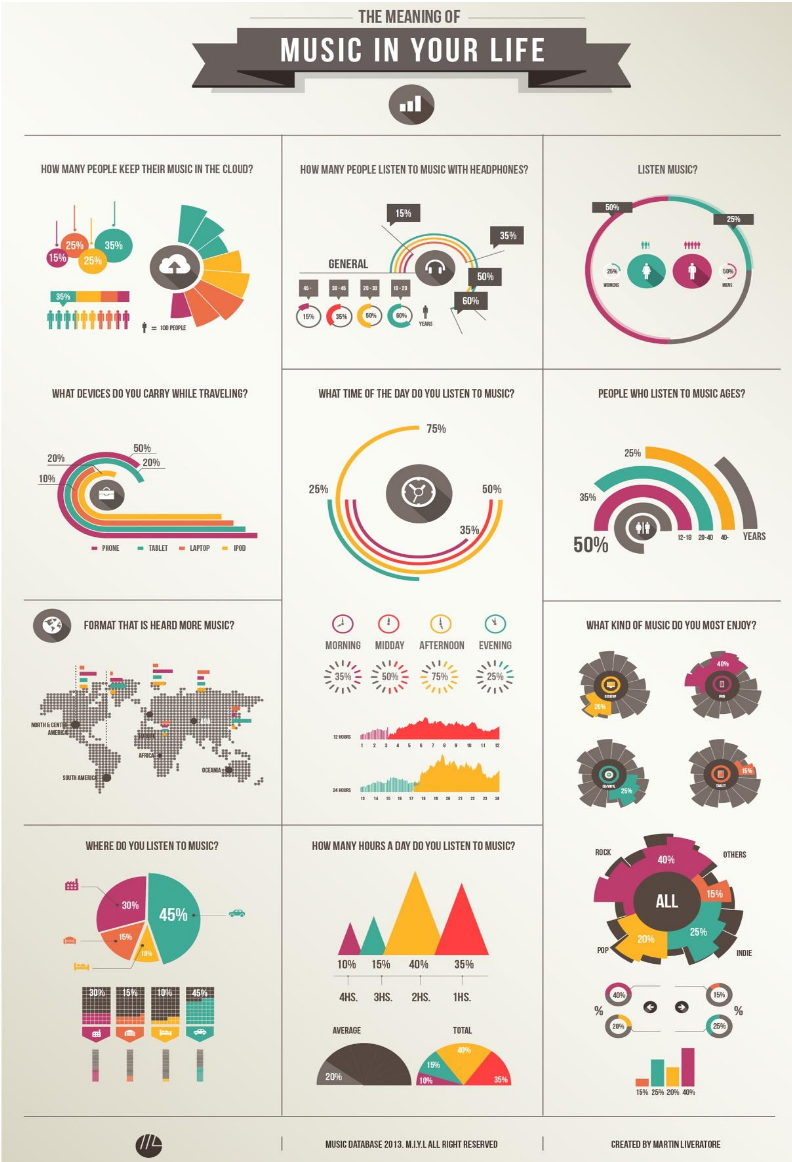


Fig 1: Instagram users and their listening habits

The infographic above was derived from a study conducted by Instagram Music in collaboration with an analytics company called Nielsen. Their aim was to track the habits and lifestyles of music listeners in the US who also are on Instagram. This was a brilliant and insightful study as we get to see how social media influences our knowledge and exposure to music. The following points can be inferred from the infographic:

1. Instagram users tend to listen to music for more hours in a week than an average US citizen, primarily because Instagram Reels has a huge impact on how we perceive music. There are a lot of songs that are often used in trends that tend to blow up and gain traction. So, using this to their advantage, a lot of artists form trends on their own songs that eventually gain popularity among the intended audience. Instagram Reels has the potential to make particular audio tracks go viral in no time, considering the amount of traffic Reels gets.
2. 42% of listeners feel that a band's live presence plays a huge role in their popularity among the public. Artist managers often use this tactic to formulate a plan for an album tour, selecting locations that are expected to have the most reception and positive response. For example, if Lil Nas X has fans mostly based in Nashville and Texas, he is more likely to tour there in order to perform sold- out shows and gain popularity.
3. The red horizontal bar chart shows what genres Instagram users are listening to. A major chunk of the population is more invested in pop and rap, with a staggering 41% of the time being spent on pop as can be seen from the graph. This makes Instagram Reels a brilliant platform to market pop music that can be used as audios for trendy videos and hence, from a marketing viewpoint, there is a higher chance of gaining traction through Instagram for a rap artist than an indie artist.
4. However, this infographic only tends to a very specific aspect of social media marketing. Instagram Music is one of the biggest audio and video social media platforms, but there are a lot of things that we don't yet know about this infographic and its data. Firstly, we have no information about the quantity of data selected and how it was collected in the first place. It may be that the data is extremely biased towards teenagers since Instagram Music was primarily used by teenagers and young adults at the time of its release. There is also no information on the location of the audience since, again, Instagram Music was exclusive to the United States at the time of release. There might be entities in this dataset that are from different age groups or different locations but we can't know that for sure.

In conclusion, does this information help a manager in formulating a path to success? That depends on how one chooses to use this data. For instance, Drake, a popular rap and R&B artist, has a very prominent following on Instagram, with his music being a popular hit on Instagram Reels and TikTok, owing to the danceability of his songs. So, for an upcoming artist with similar music to Drake (high danceability factor), Instagram would probably be a good platform to market their music on.



An infographic created by Martin Liveratore analyses the demographic and time-related factors of users' listening habits. Going through a few aspects of the infographic, the following observations are found:

1. Under the "What time of the day do you listen to music?", we can see that there is a much higher chance of someone shuffling their Spotify playlist in the afternoon than during late evening or early morning. This makes a lot of sense since the majority of the population surveyed probably belongs to the working population who are unlikely to listen to music as soon as they are back home from work or when they wake up on a tiring Monday morning.
2. "Where do you listen to music?" is a good indicator of whether a person is commuting while jamming to songs or working. This is a great measure as one would most likely not listen to

something on the heavier side (e.g. metalcore, dubstep) while driving nor would they indulge in top pop hits or rap music while working.

3. A genre preference distribution in the form of a modified pi chart and a bar chart tells us that rock music seems to be a popular choice among listeners in the US (considering all ages).
4. Even though the visualization looks impressive, the graphs are not the most easy-to-read. For example in the “People who listen to music ages?” section, there are 5 curves but only 3 categories, and even though a second look tells us that the first and the last curve don’t actually mean anything, it is still a second look that gives us that information, so it’s not very intuitive. Even in the “Where do you listen to music?” section, there are 3 graphs that give us the same information. This is again, a very unnecessary utilization of extra graphs. This could further confuse viewers as they might think that all 3 graphs portray different sets of information.

All in all, a thorough look at this graph tells us a lot about the behavior of an average listener of music in the US. In the music marketing world, the above information can be used for album releases. For instance, if there is a higher chance of listeners listening to rap at 6 PM, there is a good chance that a rap album release at around 6 PM would gain instant traction at a high volume, thus gaining a high spot in the “Trending” lists on major platforms.

Objectives and Contribution:

In this project, I aim to use modern visualization techniques to analyze how users listen to music. With this, I also aim to look at parameters such as location and time of day as well to assess an artist’s popularity and get to know their audience better. This can be achieved through the following methods:

1. Using a choropleth world map to visualize an artist’s outreach in every country. The choropleth aims to show, using colors, how many listeners listen to that particular artist in their country. This will help us get a better understanding of which artists tend to gain more international recognition and which ones are more popular in particular parts of the world.
2. Using a bar chart to visualize the time at which people listen to a particular artist. Even though this sounds very simple, it is essential in understanding the emotional responses of the public. For example, a person might be more inclined to listen to sad and melancholic music at 3 AM as it resonates with the typical physical and mental energy level during that time.

PROCESS

Data Sources

There are four datasets that are being used and all of them have been retrieved from Kaggle. The data has been collected using the Last.FM API and downloaded from Kaggle. For context, Last.FM is a service that tracks the listening activity of the user through their linked streaming service accounts. Last.FM also gives weekly reports on the users' listening patterns and provides song recommendations accordingly as well.

Dataset 1:

The first dataset shows users and their basic demographic information from Last.FM. The columns used here are:

User id, gender, age, country, registered date, country code (manually added)

Table 1:

	user_id	gender	age	country	registered	country_code
1	user_000001	m	NaN	Japan	Aug 13, 2006	JP
2	user_000002	f	NaN	Peru	Feb 24, 2006	PE
3	user_000003	m	22	United States	Oct 30, 2005	US
4	user_000005	m	NaN	Bulgaria	Jun 29, 2006	BG
5	user_000006	NaN	24	Russian Federation	May 18, 2006	RU
6	user_000007	f	NaN	United States	Jan 22, 2006	US
7	user_000008	m	23	Slovakia	Sep 28, 2006	SK
8	user_000009	f	19	United States	Jan 13, 2007	US
9	user_000010	m	19	Poland	May 4, 2006	PL
10	user_000011	m	21	Finland	Sep 8, 2005	FI
11	user_000012	f	28	United States	Mar 30, 2005	US
12	user_000013	f	25	Romania	Sep 25, 2006	RO
13	user_000015	NaN	21	Cote D'Ivoire	Oct 3, 2006	None
14	user_000016	m	NaN	United Kingdom	Aug 5, 2005	GB
15	user_000017	m	22	Morocco	Aug 27, 2007	MA
16	user_000018	NaN	22	United Kingdom	Aug 26, 2005	GB
17	user_000019	f	29	Mexico	Nov 10, 2005	MX

The second dataset also gives user details and their demographic information, but this one takes user SHA instead of user ID. This is done because the datasets used after this have either user ID or user SHA as their main ID. The columns used here are:

User SHA, gender, age, country, registered date, country code (manually added)

Table 2:

	user_sha	gender	age	country	registered	country_code
0	00000c289a1829a808ac09c00daf10bc3c4e223b	f	22.0	Germany	Feb 1, 2007	DE
1	00001411dc427966b17297bf4d69e7e193135d89	f	NaN	Canada	Dec 4, 2007	CA
2	00004d2ac9316e22dc007ab2243d6fcb239e707d	NaN	NaN	Germany	Sep 1, 2006	DE
3	000063d3fe1cf2ba248b9e3c3f0334845a27a6bf	m	19.0	Mexico	Apr 28, 2008	MX
4	00007a47085b9aab8af55f52ec8846ac479ac4fe	m	28.0	United States	Jan 27, 2006	US
5	0000c176103e538d5c9828e695fed4f7ae42dd01	m	20.0	United Kingdom	Jan 14, 2006	GB
6	0000ee7dd906373efa37f4e1185bfe1e3f8695ae	m	17.0	Finland	Nov 17, 2007	FI
7	0000ef373bbd0d89ce796abae961f2705e8c1faf	f	22.0	Poland	May 23, 2007	PL
8	0000f687d4fe9c1ed49620fbc5ed5b0d7798ea20	f	24.0	Spain	Nov 4, 2008	ES
9	0001399387da41d557219578fb08b12afa25ab67	m	NaN	Ukraine	Aug 17, 2008	UA
10	000163263d2a41a3966a3746855b8b75b7d7aa83	m	27.0	Sweden	Jan 5, 2007	SE
11	0001a57568309b287363e72dc682e9a170ba6dc2	NaN	23.0	United States	May 12, 2007	US
12	0001a88a7092846abb1b70dbcced05f914976371	NaN	NaN	Japan	Oct 13, 2008	JP
13	0001bd96207f323b53652bf400702719ad456d3c	f	18.0	Finland	Mar 10, 2008	FI
14	000215d3060a5b0ab7b3c415d49ec579100d4c87	f	NaN	Australia	Jul 26, 2008	AU
15	00024b5b85c40f990c28644d53257819980bf6bb	m	23.0	United States	Aug 19, 2005	US
16	00026e8fc41980c9605eac741cd97b8216d2dbbd	m	30.0	Portugal	Apr 22, 2007	PT

The third dataset provides a detailed view of the listening activity of all users. The mapping has been done here using user ID. The following columns have been used:

User ID, timestamp, artist ID, artist name, track ID, track name

Table 3:

	user_id	timestamp	artist_id	artist_name	track_id	track_name
0	user_000001	2009-05-04T23:08:57Z	f1b1cf71-bd35-4e99-8624-24a6e15f133a	Deep Dish	NaN	Fuck Me Im Famous (Pacha Ibiza)-09-28-2007
1	user_000001	2009-05-04T13:54:10Z	a7f7df4a-77d8-4f12-8acd-5c60c93f4de8	坂本龍一	NaN	Composition 0919 (Live_2009_4_15)
2	user_000001	2009-05-04T13:52:04Z	a7f7df4a-77d8-4f12-8acd-5c60c93f4de8	坂本龍一	NaN	Mc2 (Live_2009_4_15)
3	user_000001	2009-05-04T13:42:52Z	a7f7df4a-77d8-4f12-8acd-5c60c93f4de8	坂本龍一	NaN	Hibari (Live_2009_4_15)
4	user_000001	2009-05-04T13:42:11Z	a7f7df4a-77d8-4f12-8acd-5c60c93f4de8	坂本龍一	NaN	Mc1 (Live_2009_4_15)
5	user_000001	2009-05-04T13:38:31Z	a7f7df4a-77d8-4f12-8acd-5c60c93f4de8	坂本龍一	NaN	To Stanford (Live_2009_4_15)
6	user_000001	2009-05-04T13:33:28Z	a7f7df4a-77d8-4f12-8acd-5c60c93f4de8	坂本龍一	NaN	Improvisation (Live_2009_4_15)
7	user_000001	2009-05-04T13:23:45Z	a7f7df4a-77d8-4f12-8acd-5c60c93f4de8	坂本龍一	NaN	Glacier (Live_2009_4_15)
8	user_000001	2009-05-04T13:19:22Z	a7f7df4a-77d8-4f12-8acd-5c60c93f4de8	坂本龍一	NaN	Parolibre (Live_2009_4_15)
9	user_000001	2009-05-04T13:13:38Z	a7f7df4a-77d8-4f12-8acd-5c60c93f4de8	坂本龍一	NaN	Bibo No Aozora (Live_2009_4_15)
10	user_000001	2009-05-04T13:06:09Z	a7f7df4a-77d8-4f12-8acd-5c60c93f4de8	坂本龍一	f7c1f8f8-b935-45ed-8fc8-7def69d92a10	The Last Emperor (Theme)
11	user_000001	2009-05-04T13:00:48Z	a7f7df4a-77d8-4f12-8acd-5c60c93f4de8	坂本龍一	NaN	Happyend (Live_2009_4_15)
12	user_000001	2009-05-04T12:55:34Z	a7f7df4a-77d8-4f12-8acd-5c60c93f4de8	坂本龍一	475d4e50-cebb-4cd0-8cd4-c3df97987962	Tibetan Dance (Version)
13	user_000001	2009-05-04T12:51:26Z	a7f7df4a-77d8-4f12-8acd-5c60c93f4de8	坂本龍一	NaN	Behind The Mask (Live_2009_4_15)
14	user_000001	2009-05-03T15:48:25Z	ba2f4f3b-0293-4bc8-bb94-2f73b5207343	Underworld	dc394163-2b78-4b56-94e4-658597a29ef8	Boy, Boy, Boy (Switch Remix)
15	user_000001	2009-05-03T15:37:56Z	ba2f4f3b-0293-4bc8-bb94-2f73b5207343	Underworld	340d9a0b-9a43-4098-b116-9f79811bd508	Crocodile (Innervisions Orchestra Mix)

The fourth dataset being used here is one that gives the number of plays of an artist by a particular user. The mapping here has been done using user SHA instead of user ID. The columns being used here are:

User SHA, artist ID, artist name, total number of plays

Table 4:

	user_sha	artist_id	artist_name	plays
0	00000c289a1829a808ac09c00daf10bc3c4e223b	3bd73256-3905-4f3a-97e2-8b341527f805	betty blowtorch	2137
1	00000c289a1829a808ac09c00daf10bc3c4e223b	f2fb0ff0-5679-42ec-a55c-15109ce6e320	die Ärzte	1099
2	00000c289a1829a808ac09c00daf10bc3c4e223b	b3ae82c2-e60b-4551-a76d-6620f1b456aa	melissa etheridge	897
3	00000c289a1829a808ac09c00daf10bc3c4e223b	3d6bbeb7-f90e-4d10-b440-e153c0d10b53	elvenking	717
4	00000c289a1829a808ac09c00daf10bc3c4e223b	bbd2ffd7-17f4-4506-8572-c1ea58c3f9a8	juliette & the licks	706
5	00000c289a1829a808ac09c00daf10bc3c4e223b	8bfac288-ccc5-448d-9573-c33ea2aa5c30	red hot chili peppers	691
6	00000c289a1829a808ac09c00daf10bc3c4e223b	6531c8b1-76ea-4141-b270-eb1ac5b41375	magica	545
7	00000c289a1829a808ac09c00daf10bc3c4e223b	21f3573f-10cf-44b3-aeaa-26cccd8448b5	the black dahlia murder	507
8	00000c289a1829a808ac09c00daf10bc3c4e223b	c5db90c4-580d-4f33-b364-fbaa5a3a58b5	the murmurs	424
9	00000c289a1829a808ac09c00daf10bc3c4e223b	0639533a-0402-40ba-b6e0-18b067198b73	lunachicks	403
10	00000c289a1829a808ac09c00daf10bc3c4e223b	a342964d-ca53-4e54-96dc-e8501851e77f	walls of jericho	393
11	00000c289a1829a808ac09c00daf10bc3c4e223b	f779ed95-66c8-4493-9f46-3967eba785a8	letzte instanz	387
12	00000c289a1829a808ac09c00daf10bc3c4e223b	7b885d42-3c41-4f43-9944-a5855ec5155e	goldfrapp	361
13	00000c289a1829a808ac09c00daf10bc3c4e223b	e000d76b-afff-4286-a5fd-7ea6405cb80f	horrorpops	358
14	00000c289a1829a808ac09c00daf10bc3c4e223b	adf334c2-9186-48cd-afcc-c748dd47e9b4	the butchies	329
15	00000c289a1829a808ac09c00daf10bc3c4e223b	7e870dd5-2667-454b-9fcf-a132dd8071f1	jack off jill	316
16	00000c289a1829a808ac09c00daf10bc3c4e223b	41593aa1-dda6-4a5a-af87-0eebf93e3775	babes in toyland	310
17	00000c289a1829a808ac09c00daf10bc3c4e223b	e8374874-4178-4869-b92e-fef6bf30dc04	dropkick murphys	302
18	00000c289a1829a808ac09c00daf10bc3c4e223b	295a3ae3-9e81-4cff-a36f-8d48b8fb4dcf	all:my:faults	288
19	00000c289a1829a808ac09c00daf10bc3c4e223b	2d67239c-aa40-4ad5-a807-9052b66857a6	le tigre	281
20	00000c289a1829a808ac09c00daf10bc3c4e223b	ff7f80cd-05c2-4068-a00e-fbfbd453d049	schandmaul	244
21	00000c289a1829a808ac09c00daf10bc3c4e223b	83998f9c-846b-4294-aede-d7735531c901	edguy	232
22	00000c289a1829a808ac09c00daf10bc3c4e223b	fb01635c-51fc-4cad-b71f-62e18bb3433b	maximum the hormone	231
23	00000c289a1829a808ac09c00daf10bc3c4e223b	3f20aae7-694e-4a37-9b82-02029cf8cd4c	all ends	229

Please note that even though NaN values are present in the screenshots, they have been removed only in cases where that particular column was required.

Using these 4 root datasets, we have created more datasets using joins and filters. These datasets will further help us create visualizations that require cross-mapped data.

The first joined dataset involves the second and the fourth dataset from the list above. These two datasets in particular have been joined because we need the mapping of the number of plays of an artist in a specific country. This has been done by joining both tables on the basis of “user_sha” and the country of the user has been included in the new dataset which already contains all the columns from the “artist_plays” dataset.

Table 5:

	user_sha	artist_name	plays	country	country_code
0	00000c289a1829a808ac09c00daf10bc3c4e223b	betty blowtorch	2137	Germany	DE
1	00000c289a1829a808ac09c00daf10bc3c4e223b	die Ärzte	1099	Germany	DE
2	00000c289a1829a808ac09c00daf10bc3c4e223b	melissa etheridge	897	Germany	DE
3	00000c289a1829a808ac09c00daf10bc3c4e223b	elvenking	717	Germany	DE
4	00000c289a1829a808ac09c00daf10bc3c4e223b	juliette & the licks	706	Germany	DE
5	00000c289a1829a808ac09c00daf10bc3c4e223b	red hot chili peppers	691	Germany	DE
6	00000c289a1829a808ac09c00daf10bc3c4e223b	magica	545	Germany	DE
7	00000c289a1829a808ac09c00daf10bc3c4e223b	the black dahlia murder	507	Germany	DE
8	00000c289a1829a808ac09c00daf10bc3c4e223b	the murmurs	424	Germany	DE
9	00000c289a1829a808ac09c00daf10bc3c4e223b	lunachicks	403	Germany	DE
10	00000c289a1829a808ac09c00daf10bc3c4e223b	walls of jericho	393	Germany	DE
11	00000c289a1829a808ac09c00daf10bc3c4e223b	letzte instanz	387	Germany	DE
12	00000c289a1829a808ac09c00daf10bc3c4e223b	goldfrapp	361	Germany	DE
13	00000c289a1829a808ac09c00daf10bc3c4e223b	horrorpops	358	Germany	DE
14	00000c289a1829a808ac09c00daf10bc3c4e223b	the butchies	329	Germany	DE
15	00000c289a1829a808ac09c00daf10bc3c4e223b	jack off jill	316	Germany	DE
16	00000c289a1829a808ac09c00daf10bc3c4e223b	babes in toyland	310	Germany	DE
17	00000c289a1829a808ac09c00daf10bc3c4e223b	dropkick murphys	302	Germany	DE
18	00000c289a1829a808ac09c00daf10bc3c4e223b	all:my:faults	288	Germany	DE
19	00000c289a1829a808ac09c00daf10bc3c4e223b	le tigre	281	Germany	DE

This dataframe has been further grouped on the basis of country and artist to calculate the total number of plays which includes all users from that particular country. This has been done so that there are only two variables to plot on the choropleth at the final stage i.e. country name and total number of plays.

The next dataframe attempts to map an artist to when their song was played during the day, more specifically in which hour. We are doing this to visualize patterns in a “listening clock” manner to see emotional responses to an artist’s music. The table looks like this:

Table 6:

	user_id	timestamp	artist_id	artist_name	track_id	track_name	gender	age	country	registered	country_code	hour
48885	user_000002	2006-12-26 20:56:45+00:00	83d91898-7763-47d7-b03b-b92132375c47	Pink Floyd	2d1201cf-59bb-4ffa-9f52-f5b3afa13346	Us And Them	f	NaN	Peru	Feb 24, 2006	PE	20
48886	user_000002	2006-12-26 20:50:23+00:00	83d91898-7763-47d7-b03b-b92132375c47	Pink Floyd	7fef22bd-76aa-4803-b56b-93a5d6e70662	Money	f	NaN	Peru	Feb 24, 2006	PE	20
48887	user_000002	2006-12-26 20:48:23+00:00	83d91898-7763-47d7-b03b-b92132375c47	Pink Floyd	7fef22bd-76aa-4803-b56b-93a5d6e70662	Money	f	NaN	Peru	Feb 24, 2006	PE	20
48888	user_000002	2006-12-26 20:45:36+00:00	83d91898-7763-47d7-b03b-b92132375c47	Pink Floyd	4c3dd9dd-04b1-48d0-b7d8-45974080e9da	The Great Gig In The Sky	f	NaN	Peru	Feb 24, 2006	PE	20
48889	user_000002	2006-12-26 20:38:31+00:00	83d91898-7763-47d7-b03b-b92132375c47	Pink Floyd	0e11c0fd-a1da-4b88-a438-7ef55c5809ec	Time	f	NaN	Peru	Feb 24, 2006	PE	20
48890	user_000002	2006-12-26 20:36:31+00:00	83d91898-7763-47d7-b03b-b92132375c47	Pink Floyd	0e11c0fd-a1da-4b88-a438-7ef55c5809ec	Time	f	NaN	Peru	Feb 24, 2006	PE	20
48891	user_000002	2006-12-26 20:34:44+00:00	83d91898-7763-47d7-b03b-b92132375c47	Pink Floyd	19084069-642f-465f-9127-f71bcd800a05	On The Run	f	NaN	Peru	Feb 24, 2006	PE	20
48892	user_000002	2006-12-26 20:30:57+00:00	83d91898-7763-47d7-b03b-b92132375c47	Pink Floyd	d59cccb9-72da-45e6-be9f-84808deb8d33	Speak To Me	f	NaN	Peru	Feb 24, 2006	PE	20
48893	user_000002	2006-12-26 20:28:58+00:00	83d91898-7763-47d7-b03b-b92132375c47	Pink Floyd	d59cccb9-72da-45e6-be9f-84808deb8d33	Speak To Me	f	NaN	Peru	Feb 24, 2006	PE	20
60969	user_000002	2006-08-29 18:24:25+00:00	83d91898-7763-47d7-b03b-b92132375c47	Pink Floyd	feecff58-8ee2-4a7f-ac23-dc8ce7925286	Wish You Were Here	f	NaN	Peru	Feb 24, 2006	PE	18
60970	user_000002	2006-08-29 18:21:29+00:00	83d91898-7763-47d7-b03b-b92132375c47	Pink Floyd	c46641c1-fdbc-4401-8b1d-23a4062a207a	Eclipse	f	NaN	Peru	Feb 24, 2006	PE	18
60971	user_000002	2006-08-29 18:18:37+00:00	83d91898-7763-47d7-b03b-b92132375c47	Pink Floyd	NaN	Braindamage	f	NaN	Peru	Feb 24, 2006	PE	18

The final ready-to-plot table that we get looks something like this:

	hour	Counts
0	0	476
1	1	499
2	2	589
3	3	727
4	4	639
5	5	556
6	6	592
7	7	461
8	8	385
9	9	352
10	10	284
11	11	306
12	12	292
13	13	224
14	14	298
15	15	371
16	16	384
17	17	422
18	18	535
19	19	490

Finally, the genre of tracks has been found using the Last.FM API and a call to the track.getTopTags() function through the API. We have taken a sample of 10000 songs to see which genre is heard more in which country and during which time of the day. The table looks something like this:

Table 7:

	user_id	timestamp	artist_id	artist_name	track_id	track_name	gender	age	country	registered	country_code	hour	genre
0	user_000934	2009-02-18 13:25:38+00:00	f37c537b-3557-4031-bfd6-ab63ced32854	10Cc	1e584d84-d0c8-4448-9b06-114717044c2e	I'M Not In Love	m	NaN	Northern Mariana Islands	May 13, 2006	MP	13	70s
1	user_000503	2006-06-11 01:09:57+00:00	013fa897-86db-41d3-8e9f-386c8a34f4e6	Morrissey	327063f8-426f-40a3-aa02-96af1da0982a	In The Future When All'S Well	f	NaN	Austria	Dec 3, 2005	AT	1	alternative
2	user_000170	2006-04-04 22:37:33+00:00	9bffb20c-dd17-4895-9fd1-4e73e888d799	モーニング娘。	b82210af-6f54-4587-9093-ce92721cd7fd	Morning Coffee	f	29	Canada	Apr 27, 2005	CA	22	JPop
3	user_000540	2008-08-01 13:37:10+00:00	49df3848-1eda-44bb-962f-a9f29fffe204	4C3	c88ce0a7-bebe-4ad1-abed-929c6a89f81a	Last Skyline	f	NaN	Spain	May 24, 2007	ES	13	NaN
4	user_000834	2007-09-23 18:26:58+00:00	847e8284-8582-4b0e-9c26-b042a4f49e57	Placebo	0792b3ec-8e43-46d8-8933-b8ea9d45c5d4	Come Home	f	NaN	Italy	Apr 7, 2006	IT	18	alternative rock
5	user_000834	2007-07-21 15:07:57+00:00	7e870dd5-2667-454b-9fcf-a132dd8071f1	Jack Off Jill	d4dedca3-762a-4106-bf28-35f4d63508e1	Strawberry Gashes	f	NaN	Italy	Apr 7, 2006	IT	15	alternative rock
6	user_000270	2006-12-23 05:05:22+00:00	af065f9a-2c64-4e3e-84b1-1d20d6b13fb3	Chronic Future	cfd7717e-3843-482e-a749-d35ebc1112fd	Time And Time Again	f	20	United States	Jan 3, 2006	US	5	punk rock
7	user_000511	2006-12-17 03:20:21+00:00	6c8b9855-ba8b-48f9-ac1d-42167f7f7b18	The Shins	5de7df64-b83a-4f05-9c3c-bfeb0a0bf0d2	Australia	f	NaN	United States	Dec 14, 2005	US	3	indie
8	user_000768	2008-04-10 16:37:46+00:00	83e59f23-3b0b-4304-834d-5bcafd5df6d2	Scooter	b49f9385-7131-443b-b9de-cbd0a5fba5ff	Devil Drums	m	NaN	Romania	Jun 4, 2007	RO	16	trance
9	user_000712	2008-09-15 [illegible]	172e1f1a-504d-4488-[illegible]	Nick Cave & The Bad [illegible]	68222d24-0216-4bda-[illegible]	Red Right [illegible]	m	NaN	Poland	Mar 4, [illegible]	PL	12	rock

RESULTS AND INSIGHTS

Choropleth:

A choropleth is defined as a world map that uses colours and it's different shades to represent frequency and density. This is best to compare frequencies according to countries.

In this case, we have analyzed how many users listen to particular artists. We have taken 3 artists for our case (Coldplay, Kanye West, Betty Blowtorch). The graphs are as follows:

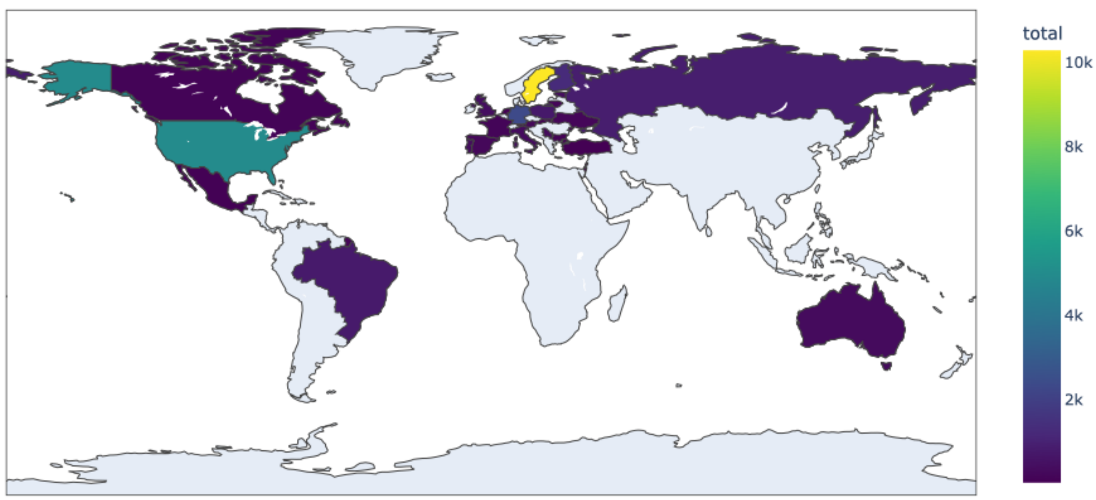


Fig 1: Number of Listeners per Country for the band Betty Blowtorch

Betty Blowtorch is an all-female American hard rock band that was formed in the 1990s in Southern California. Even though all the members were American and there was very minimal scope of international exposure, the graph above tells us otherwise. We can see that on the map, Sweden has the most number of listeners for this band! This could be due to any reason whatsoever, but the bottom line is that a band from the USA will not necessarily have listeners from only the USA or its neighbouring countries. For Betty Blowtorch's manager (if they were active in the present world of big data analytics), this could mean that Sweden is a gateway of gaining exposure in Europe and subsequently to other continents.

Similarly, here are the graphs for Coldplay and Kanye West respectively:

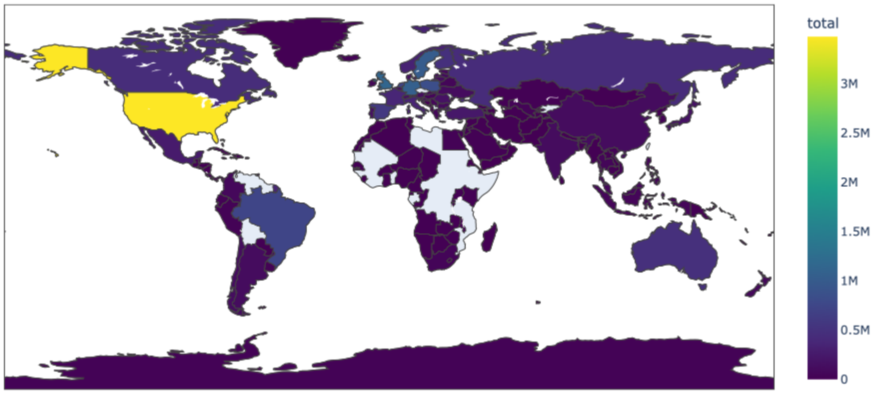


Fig 2: Number of Listeners per Country for the band Coldplay

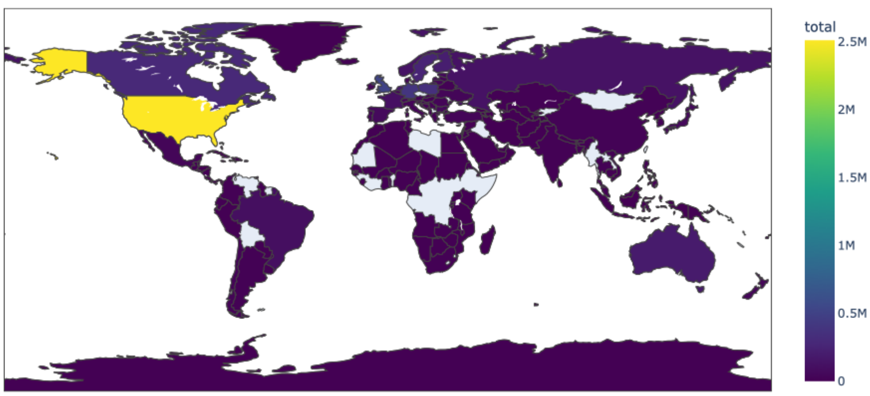


Fig 3: Number of Listeners per Country for the rapper Kanye West

We can infer a few things from the above two maps about Coldplay and Kanye West. We can see from the colors that Coldplay has more international listeners than Kanye West, particularly in the regions of Latin America, Russia, Europe and Australia. This is due to the fact that Coldplay often performs internationally and they have been doing it almost since the start of their career. This gave a huge boost to their fame and got them to where they are now. On the other hand, Kanye has more of one-directional fame mostly through the United States. Kanye's manager, please take notes!

There is a small issue that was faced here while plotting this choropleth. The default colormap as well as other colormaps did not fit well with the visualization and gave results that could confuse a lot of people, especially colorblind people. This is an example of the plot for Coldplay (refer Fig 2) but in the default colormap:

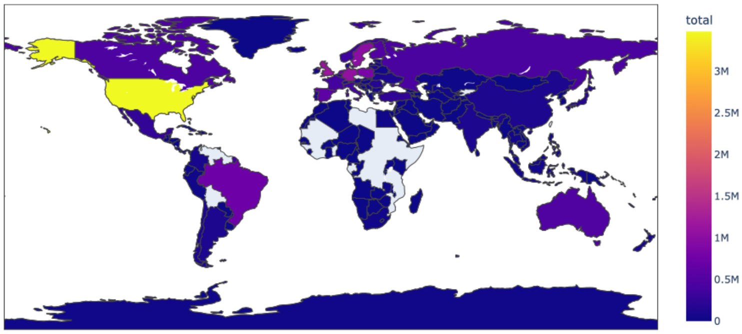


Fig 4: Number of Listeners per Country for the band Coldplay (Default Colormap)

As we see above, there is a high chance that someone could confuse the blue with the purple, and get wrong inferences. This is why using perceptually uniform colormaps is a huge necessity, and should be for all visualizations.

Listening Clock (Bar Plot):

This next visualization gets us some interesting inferences. The reason we refer to this visualization as a listening clock is because the graph shows, for a selected artist, what time of the day their music is heard the most. This might seem like very plain and intuitive information, but it has proven to be of monumental significance to artist managers in the past. Moving forward, we have essentially used a bar plot with a “colorwarm” colormap. The x and y axis are described as follows:

- The x-axis gives us the hour in which the artist was played. We have used a 24-hour clock here to avoid confusion so the x-axis ticks are essentially just integers from 0 to 23.
- The y-axis gives us the count of the number of listeners listening to the particular artist’s music at the given time. Here is a look at the graph now:

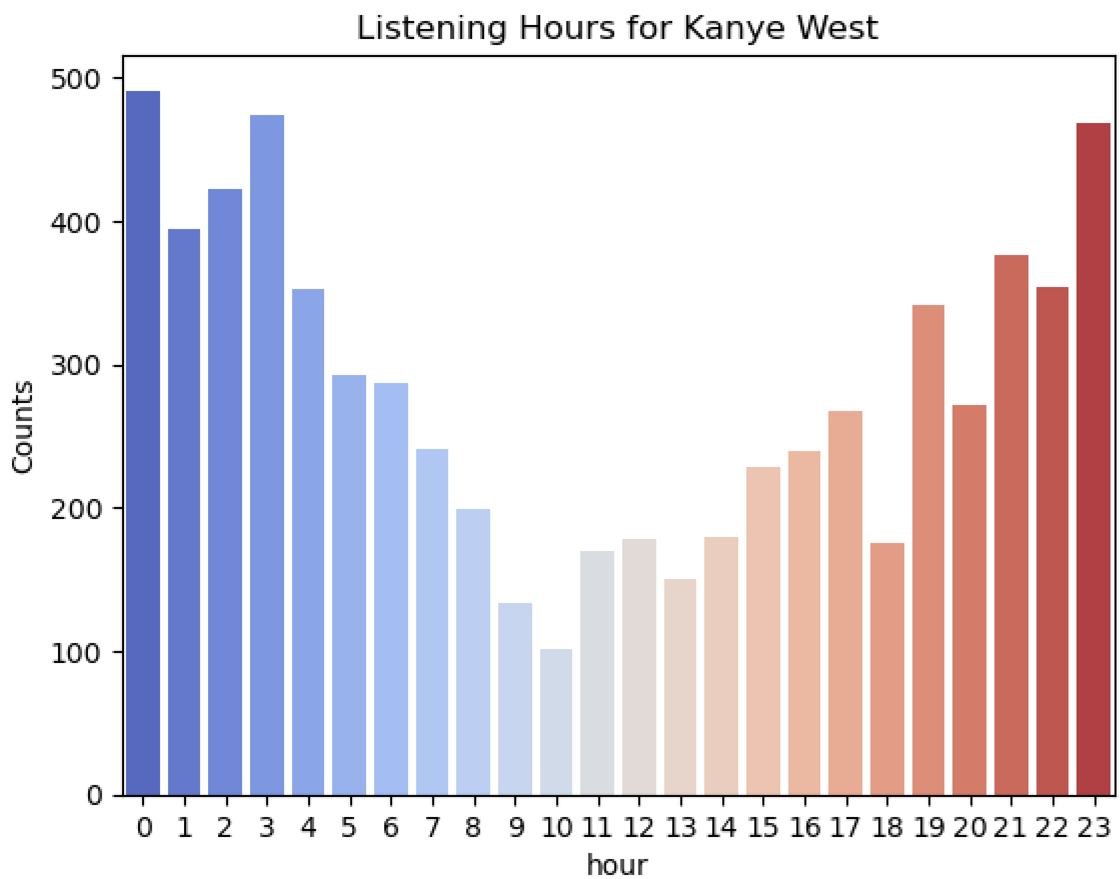


Fig 5: Listening Hours for Kanye West

As we see above, Kanye's music is played more often during the hours of 11 PM and 12 AM, with considerable peaks at 7 PM and 9 PM as well. This is pretty understandable as rap music is seen more of an "after-dark" mood of music. The peaks at 7 PM and 9 PM probably indicate streaming traffic from people commuting at the evening rush hour or commuting to a party.

On to the next band:

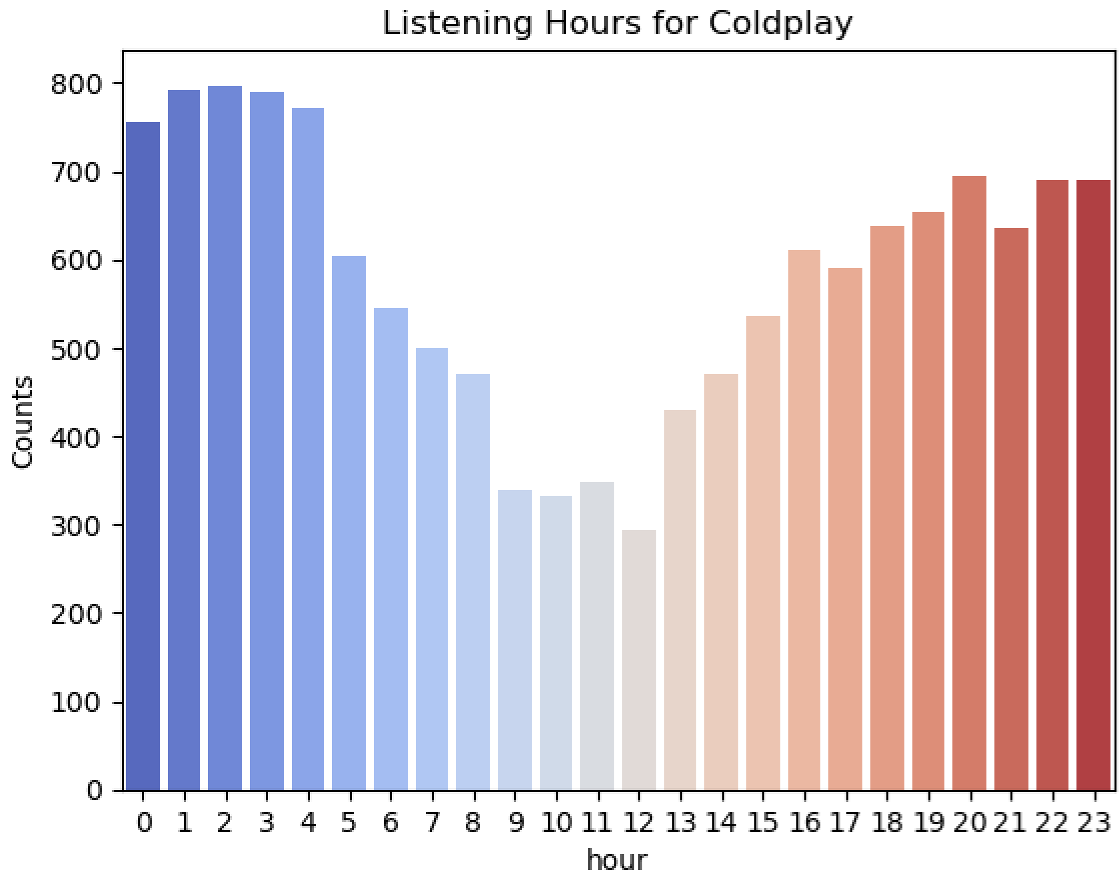


Fig 6: Listening Hours for Coldplay

Coldplay is a band known for their melodic tunes one would sway their head to, deep and raw vocals that hit every emotion one could experience and ethereal lyrics that transports one to a utopian dreamland. Doesn't sound like a very upbeat band, does it? The above graph shows us exactly that! There is a high volume of traffic for every hour, however we can see considerable bar heights for the hours from 1 AM to 5 AM. This is typically when the mind and body are close to falling asleep and the tiredness and fatigue requires music that resonates with the same, hence, Coldplay.

On to another interesting rock band:

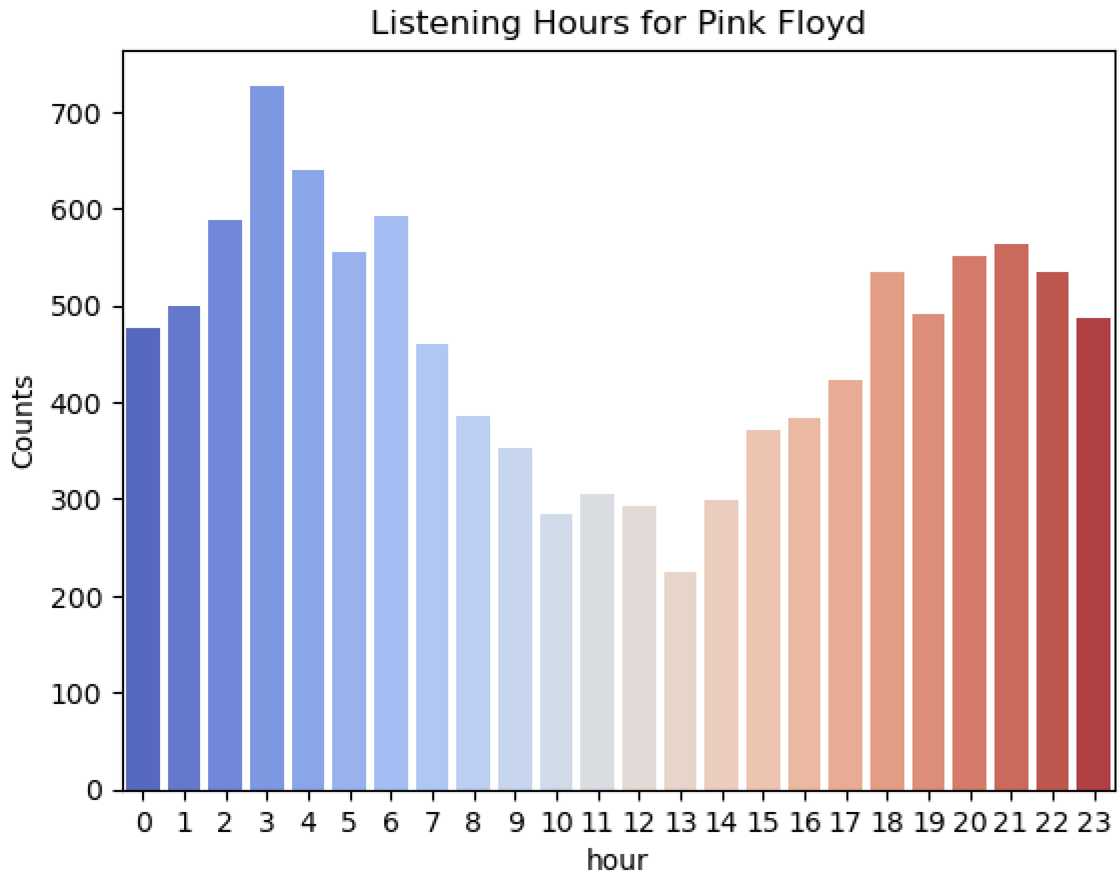


Fig 7: Listening Hours for Pink Floyd

Pink Floyd is one of the most influential bands of the 1970s classic rock era. They emerged with a completely new way of creating music that felt divine and graceful to listeners. This is also around the time when psychedelic rock started to become a thing. Coincidentally, Pink Floyd's music releases around this time started to become more slow tempoed and "trippy", as described by most drug users. This explains the massive peak of listeners at the 3-4 AM mark, owing to their slow-tempoed music.

One thing that I especially made sure of while creating these bar plots is that the y-axis remains relative to the artist and the plot. This is because using an axis that is fixed will get longer bars in some bands and shorter in others. Viewers could mistake this as the artist having a very low or a very high number of listeners overall.

DISCUSSION:

This project has been an insightful experience into a very real and very interesting problem of one of the biggest fields of entertainment in the world. There are countless ways in which we can visualize and analyze customer data like this and the marketing department of any industry is the biggest and the best example for this. From this project, I could find that even data as simple as just the listening activity of users from a streaming service tracking website can be so useful in seeing the performance of particular artists. There were even a few surprises that I found while playing around with the data and its possible visualizations. For example, Betty Blowtorch, being a primarily all-American band, had their most listeners in Sweden instead of the United States as most people would think! Even in the listening clock examples, we saw that even though Kanye West technically has more plays on his songs on streaming services than Coldplay does, it is surprising to know that Coldplay has more international recognition than Kanye! Data can be monumental in making decisions, especially for marketing and business, and this project has been a prime example of that.

CONCLUSION:

Data and its visualizations are a superb communication tool that can be used for things as essential as closing a company buyout deal or as small as convincing your dad to buy you a car! The visualizations above, that have used users' music listening activity from different streaming services collected on last.fm, have given us a lot of inferences about success rates and audience's behavior for a lot of artists. Country by country popularity for selected artists was analyzed through different choropleths, through which we found the anomaly of Betty Blowtorch. The other case was using barplots to analyze the time of day when listeners listen to a particular artist's music. This gave us the inference of how the human brain reacts and responds to different kind of music.

FUTURE WORK:

This project is just a glimpse of what big data analytics has the potential to do for the music industry. The problem is that even though this already has been implemented in the music industry, smaller artists who are just starting out do not have the opportunity to take advantage of advanced marketing techniques as only major record labels have the accessibility to do something like this for their signed artists. For the future, this can be used for up and coming indie artists, more specifically in more localized regions, so that more people have access to this kind of technology.

REFERENCES:

1. Case Study Visualization 1: https://medium.com/cuepoint/what-instagram-discovered-in-our-first-nielsen-music-study-de1a2740c005#.qgey4a9u6
2. Case Study Visualization 2: https://www.behance.net/gallery/19624167/Infography-Set-The-music-in-your-life
3. https://www.inc.com/darren-heitner/big-data-is-revolutionizing-music-industry-here-are-lessons-for-your-business.html
4. https://musicentrepreneurhq.com/main-challenges-music-marketers-facing-internet/
5. https://www.ucf.edu/pegasus/your-brain-on-music/
6. https://www.complex.com/music/2019/09/data-changing-music-industry
7. https://www.rollingstone.com/pro/features/snafu-records-launches-with-artificial-intelligence-948212/
8. https://www.wired.com/story/big-data-music/
9. https://www.uagc.edu/blog/how-does-music-affect-your-brain
10. https://pudding.cool/2021/04/music-bubble/
11. Dataset 1: https://www.kaggle.com/kerneler/starter-last-fm-2337f84b-a/data?select=userid-profile.csv
12. Dataset 2: https://www.kaggle.com/kerneler/starter-last-fm-2337f84b-a/data?select=userid-timestamp-artid-artname-traid-traname.tsv
13. https://grouplens.org/datasets/hetrec-2011/