Radio Data in a Digital Age

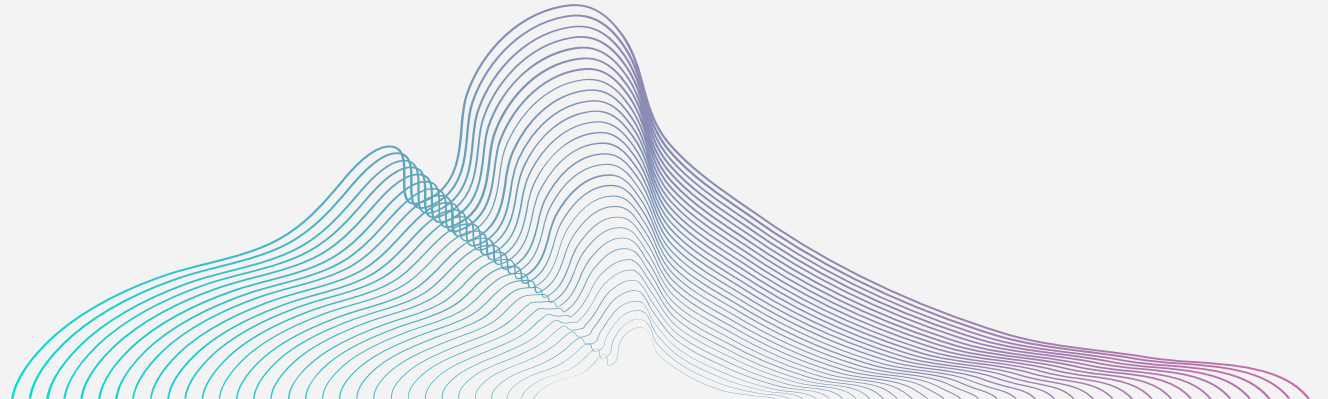
Lisa Girard
Thinkful - Data Science Flex
Capstone #1
8.17.22

About Me...

- Future Data Scientist - student in Thinkful DS Flex Bootcamp
- Digital marketing innovator and strategist with 30+ years experience in radio, TV, web, email, social media, digital advertising, SEO, mobile app, and offline digital applications
- Masters of Arts, Digital Media, University of Denver, 2005
- Communications Strategist and PR/Marketing Evangelista documentary film: *35000 Watts: The Story of College Radio*
- Former radio DJ
- Music and pop culture aficionado

Radio Data in a Digital Age

Tracking charting history on the Billboard Hot 100 chart



Billboard Hot 100

The ***Billboard Hot 100*** is one of Billboard Magazine's definitive music charts and is powered by Luminate, the entertainment industry data powerhouse.

The week's most popular songs are ranked by audio and video streaming activity on leading digital music services, radio airplay audience impressions based on monitored airplay, and sales data.

A publicly-available dataset of all "Hot 100" charts released since its inception in 1958 through 11.6.21 can be found online at kaggle.com/datasets/dhruvildave/billboard-the-hot-100-songs

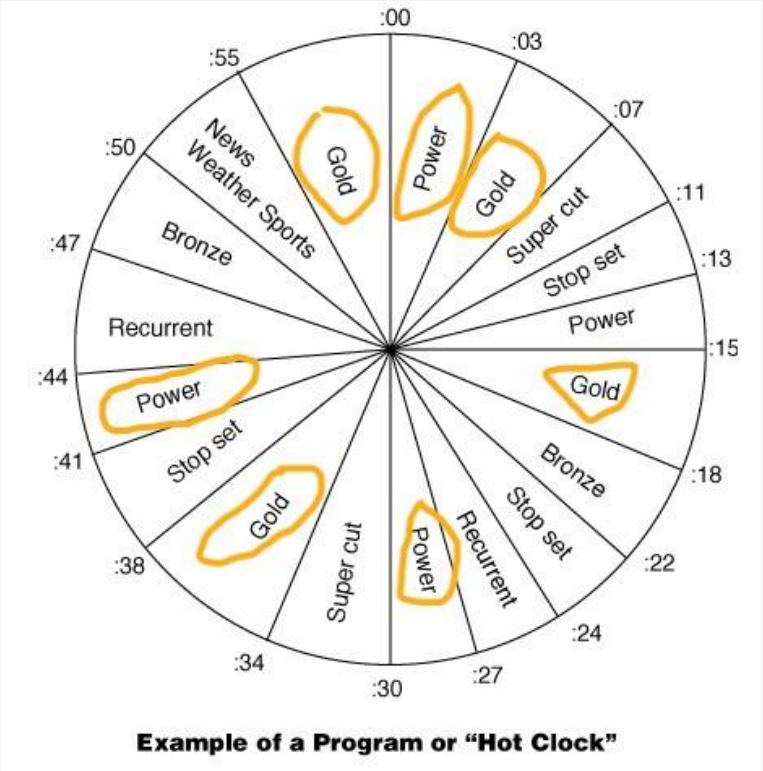
Data Exploration & Observations

Backstory:
Radio Format Clock and Power Gold Hits

The broadcast radio format is determined by a programmed “clock” where music directors, management, sales teams, and others determine when to play what kind of songs during certain hours of the day.

These ‘clocks’ often contain trending hits, past current hits, and what are known as **gold or power gold** hits.

Power gold hits are often the “Stairway to Heaven”, “Highway to Hell”, and “Brown-Eyed Girl” hits that anyone can find playing on any radio station, at any time.



Example of a Program or “Hot Clock”

Power Gold Hits of the ‘70s

Songs with an almost 50-year shelf life were charting for less than 20 weeks when they were released.

Power gold hits of the 1970s like:

- “Hotel California”
- “One of These Nights”
- “More Than A Feeling”
- “Amanda”
- “Don’t Look Back”

were charting for less than 20 weeks.

	date	rank	song	artist	last-week	peak-rank	weeks-on-board
237286	1976-05-22	100	Take It To The Limit	Eagles	82.0	4	23
231482	1977-07-02	83	Hotel California	Eagles	58.0	1	19
243145	1975-04-05	59	Best Of My Love	Eagles	38.0	1	19
240734	1975-09-20	48	One Of These Nights	Eagles	29.0	1	17
216195	1980-06-07	96	I Can't Tell You Why	Eagles	86.0	8	16
217398	1980-03-15	99	The Long Run	Eagles	84.0	8	15
218298	1980-01-12	99	Heartache Tonight	Eagles	34.0	1	15

	date	rank	song	artist	last-week	peak-rank	weeks-on-board
233741	1977-01-22	48	More Than A Feeling	Boston	29.0	5	19
181597	1987-01-24	98	Amanda	Boston	77.0	1	18
180892	1987-03-14	93	We're Ready	Boston	52.0	9	15
179794	1987-05-30	95	Can'tcha Say (You Believe In Me)/Still In Love	Boston	87.0	20	13
224394	1978-11-11	95	Don't Look Back	Boston	50.0	4	13
223197	1979-02-03	98	A Man I'll Never Be	Boston	52.0	31	12
142198	1994-08-13	99	I Need Your Love	Boston	86.0	51	10
232781	1977-04-02	82	Long Time	Boston	43.0	22	10
231464	1977-07-02	65	Peace Of Mind	Boston	38.0	38	8
221895	1979-05-05	96	Feelin' Satisfied	Boston	46.0	46	7

Power Gold Hits of the 2000s

Songs released in the last 20 years, were charting much, much longer than ‘20 weeks or less’.

Imagine Dragons maxed out at *87 weeks* on the charts from a song released in 2014.

	date	rank	song	artist	last-week	peak-rank	weeks-on-board
90147	2004-07-31	48	Someday	Nickelback	46.0	7	50
100449	2002-08-10	50	How You Remind Me	Nickelback	45.0	1	49
71948	2008-01-26	49	Rockstar	Nickelback	38.0	6	49
81147	2006-04-22	48	Photograph	Nickelback	43.0	2	33
76747	2007-02-24	48	Far Away	Nickelback	36.0	8	30
78949	2006-09-23	50	Savin' Me	Nickelback	42.0	19	28
65549	2009-04-18	50	Gotta Be Somebody	Nickelback	47.0	10	27
74947	2007-06-30	48	If Everyone Cared	Nickelback	41.0	17	22
63644	2009-08-29	45	If Today Was Your Last Day	Nickelback	32.0	19	22
100681	2002-07-27	82	Too Bad	Nickelback	77.0	42	20
58397	2010-09-04	98	This Afternoon	Nickelback	84.0	34	20
90898	2004-06-12	99	Figured You Out	Nickelback	97.0	65	19
60998	2010-03-06	99	Never Gonna Be Alone	Nickelback	80.0	58	17
89094	2004-10-16	95	Feelin' Way Too Damn Good	Nickelback	86.0	48	16
51095	2012-01-28	96	When We Stand Together	Nickelback	77.0	44	16
81396	2006-04-08	97	Animals	Nickelback	NaN	97	3
67394	2008-12-13	95	I'd Come For You	Nickelback	44.0	44	2
51788	2011-12-10	89	Lullaby	Nickelback	NaN	89	1
67795	2008-11-15	96	Something In Your Mouth	Nickelback	NaN	96	1

	date	rank	song	artist	last-week	peak-rank	weeks-on-board
39148	2014-05-10	49	Radioactive	Imagine Dragons	48.0	3	87
38346	2014-07-05	47	Demons	Imagine Dragons	45.0	6	61
18344	2018-05-05	45	Thunder	Imagine Dragons	34.0	4	52
19528	2018-02-10	29	Believer	Imagine Dragons	28.0	4	52
44348	2013-05-11	49	It's Time	Imagine Dragons	48.0	15	47
14449	2019-02-02	50	Natural	Imagine Dragons	31.0	13	27
16936	2018-08-11	37	Whatever It Takes	Imagine Dragons	34.0	12	27
34257	2015-04-18	58	I Bet My Life	Imagine Dragons	54.0	28	22
25641	2016-12-10	42	Sucker For Pain	Lil Wayne, Wiz Khalifa & Imagine Dragons With ...	37.0	15	22
38397	2014-07-05	98	On Top Of The World	Imagine Dragons	91.0	79	20
1585	2021-07-24	86	Follow You	Imagine Dragons	77.0	68	15
13184	2019-05-04	85	Bad Liar	Imagine Dragons	80.0	56	9
34888	2015-03-07	89	Shots	Imagine Dragons	98.0	75	3
17493	2018-07-07	94	Born To Be Yours	Kygo & Imagine Dragons	74.0	74	2
32076	2015-09-19	77	Roots	Imagine Dragons	NaN	77	1
20068	2018-01-06	69	Thunder/Young Dumb & Broke (Medley)	Imagine Dragons + Khalid	NaN	69	1
42277	2013-10-05	78	Monster	Imagine Dragons	NaN	78	1

Nickelback had hits released in 2002 and 2004 that charted around 50 weeks.

Hypothesis & Data Testing

I hypothesize that "music in the digital age" charted longer because of more equitable access and quantifiable measurement available with digital file sharing and streaming music services.

Therefore, I segmented the Hot 100 dataset based on time frame of pre-digital age and post-digital age.

H_0 - Songs charted from 1976-1998 did not have significantly different chart times than songs released after 1999.

H_1 - Songs charted from 1999-2021 did have significantly different chart times (longer) than songs released between 1976-1998.

Stakeholders can use this information to make data-informed decisions regarding content distribution, audience development, and marketing/revenue generation.

Stakeholders

Record Labels

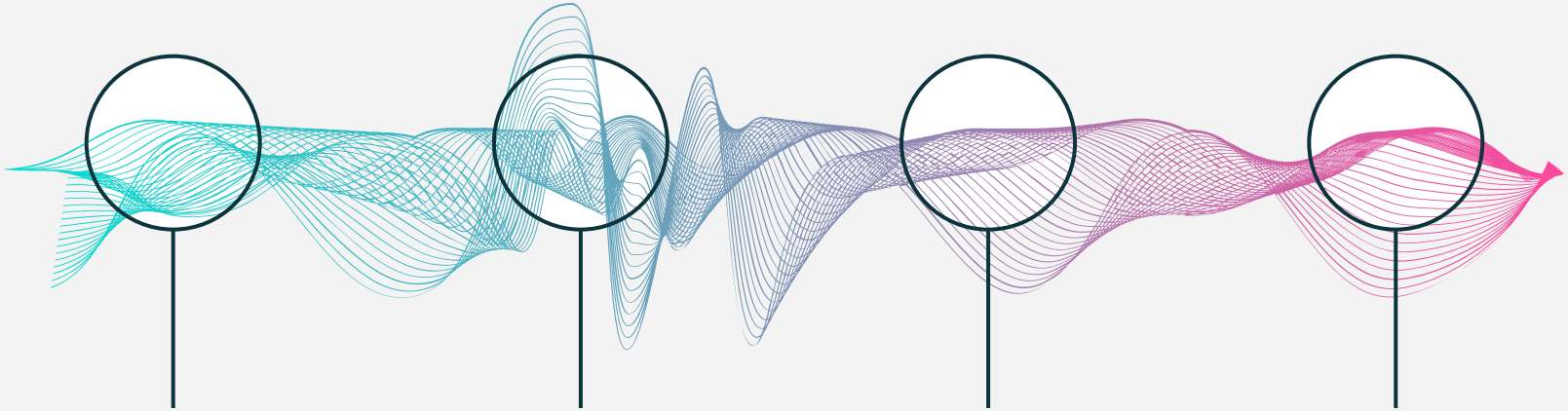
Artists

Streaming Music Services

Traditional Broadcast Radio

Satellite Radio Stations

An Evolutionary Timeline of Digital Music Content Delivery



1999

Napster launches free file-sharing program to share digital music files, denting record sales, launching platforms to allow for digital music content delivery.

2001

Napster/Rhapsody - launches first digital music streaming service based on subscription model

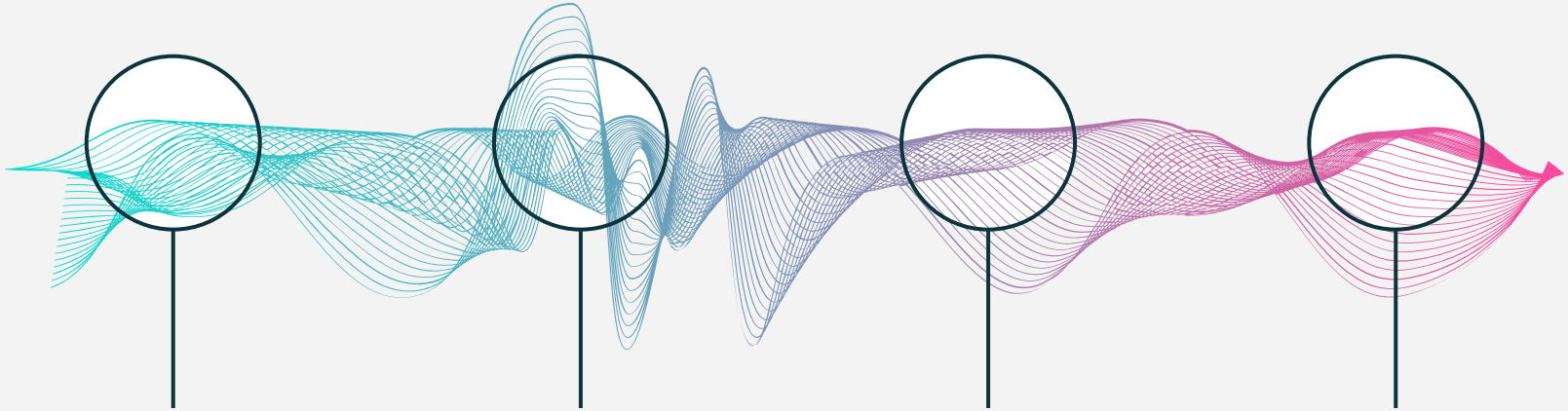
2002

LAST.fm - online radio platform, based on user's patterns (streaming service)

2003

iTunes - Apple launches digital music store for song file purchases (sales)

An Evolutionary Timeline of Digital Music Content Delivery



2005

Pandora - online radio platform, based on users' preferences and listening habits, algorithm driven (streaming service)

2007

SoundCloud launches - streaming platform geared towards unsigned artists

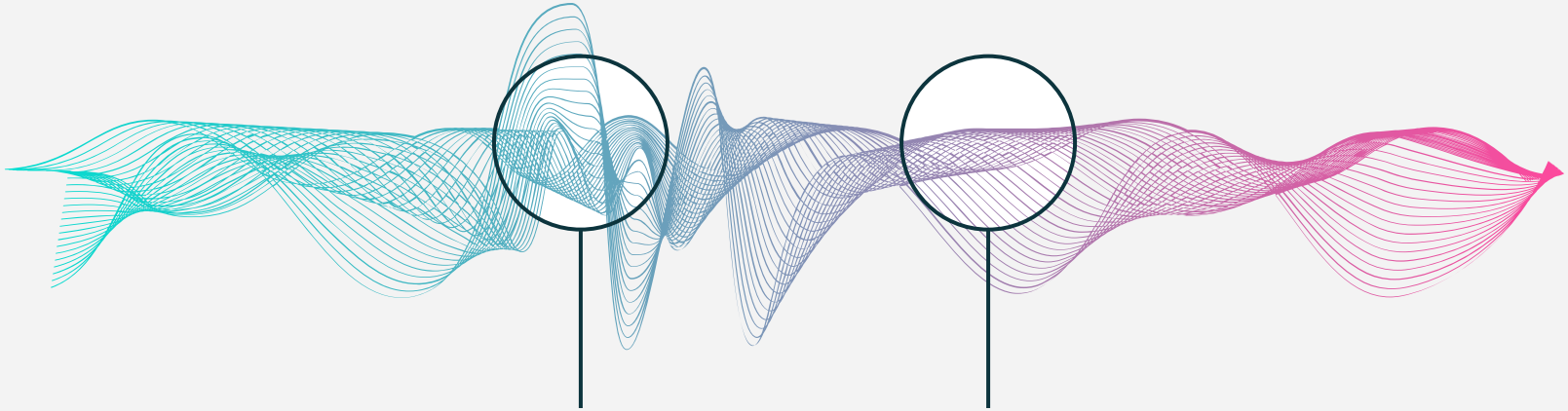
2008

Bandcamp - fileshare and streaming platform geared towards unsigned artists

2011

Spotify comes to the USA - streaming music service, with/without ads, subscription model for ad-free

An Evolutionary Timeline of Digital Music Content Delivery



2013

YouTube becomes noted music streaming provider (not just videos) per "Gangnam Style" by PSY - hits first 1B views, Billboard starts charting digital streaming services in addition to radio plays and sales

2019

Streams pass 1 trillion, Spotify hits 100M subscribers

source: .billboard.com/music/music-news/2013-year-of-streaming-8545169/

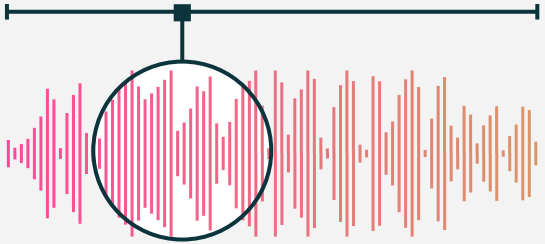
The Research

Comparing chart time (“weeks on board”) for songs released prior to 1999 to chart time for songs released after 1999.

Pre-1999

Data set: 22 years prior to 1999
Songs released from
1.1.1976 to 12.31.1998

A group



Post-1999

Data set: 22 years after to 1999
Songs released from
1.1.1999 to 11.06.2021

B group

The Research

Testing Methods and Considerations

Testing was carried out as a standard A/B test for two unrelated variables (“pre-1999” and “post-1999”)

Analyses were performed that considered:

- Sample size / Bias
- Clean data (nulls and outliers)
- Variable relationship
- Normality of distributions
- Statistical non-parametric testing
- Confidence interval

Data Wrangling: Sample Size & Bias

- Original database: 330K rows
- 2013 - 2021: 46,200 rows **
- 2005 - 2012: 36,600 rows **
- 1999 - 2021: 120K rows
- 1976 - 1998: 119K rows

** Originally considered 2013 as the digital milestone because that's when Billboard started including streaming stats. But that did not account for all digital content delivery from 1999 - 2012, nor did a timeframe of 8 years prior to 2013.

Decisions:

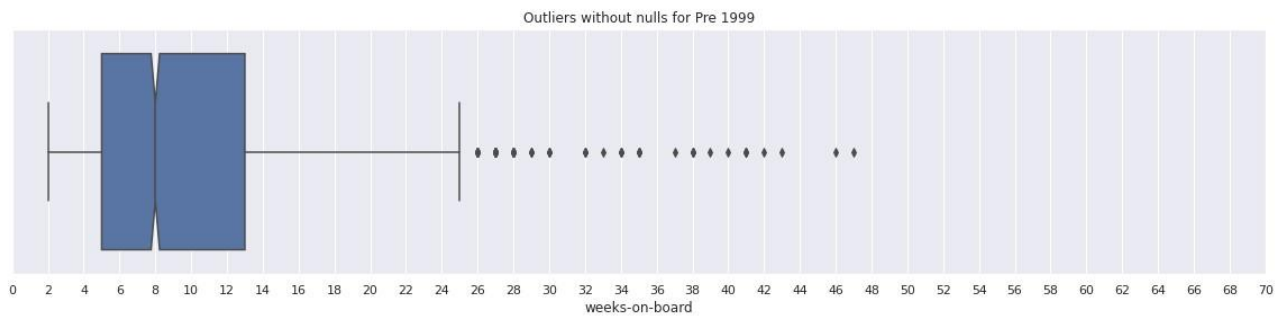
- Stick with 1999 as digital milestone year
- Set timeframes for data sets to be +/- 22 years from 1999
- Evaluate only for "weeks-on-board"
- Randomize data sets to account for smaller sample size and bias - 3000 rows to evaluate instead of 120K

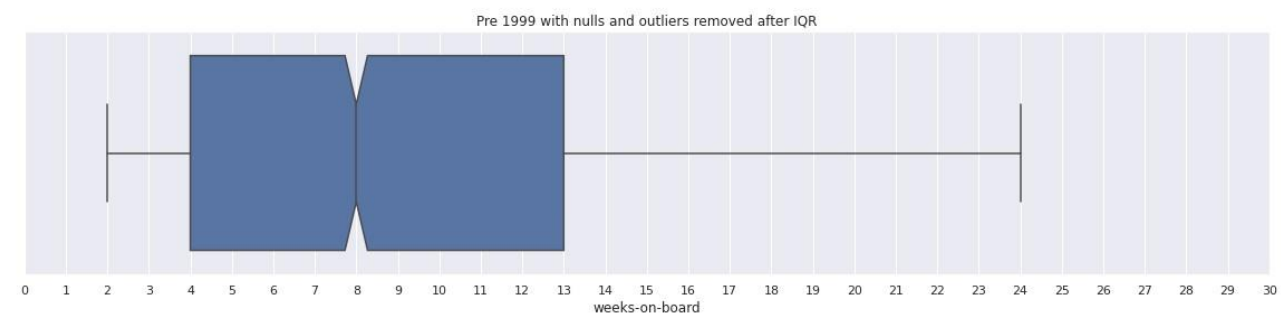
2 Variables to Test:

- Pre-1999 for "Weeks on Board" (A)
  3000 row random sample size
- Post-1999 for "Weeks on Board" (B)
  3000 row random sample size

Data Cleaning: Nulls & Outliers

- “NaN’s” (or nulls) removed from both Pre-1999 and Post-1999 data sets
- Outliers for Pre-1999 started around weeks 25-27; outliers for Post-1999 started around weeks 31-33
- Calculated IQR values to remove outliers for both data sets





Variable Relationship

- Variable relationship determines future testing methods and outcome interpretations
- Is the data within “Pre-1999 for weeks on board” related in any way to the data in “Post-1999 for weeks on board”?

Two additional tests indicated that Pre-1999 and Post-1999 are NOT related

Pre-1999

“Weeks on Board”

Post-1999

“Weeks on Board”

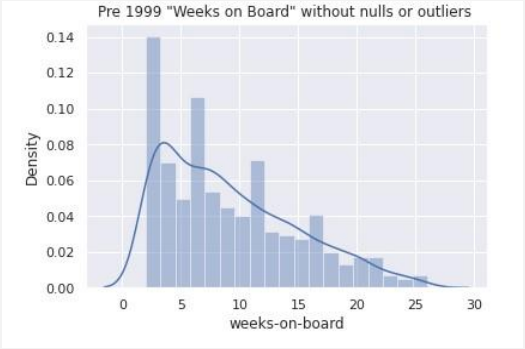
Why Be Normal?

Determining Distribution and Testing Methods for Data Analysis

The Color and Shape of: The Data

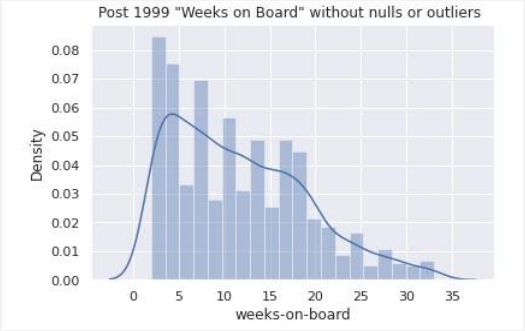
Pre & Post: stats.describe

Stats.describe on cleaned data shows as mostly normal but histogram plots show data is still right skewed and *not normally distributed*



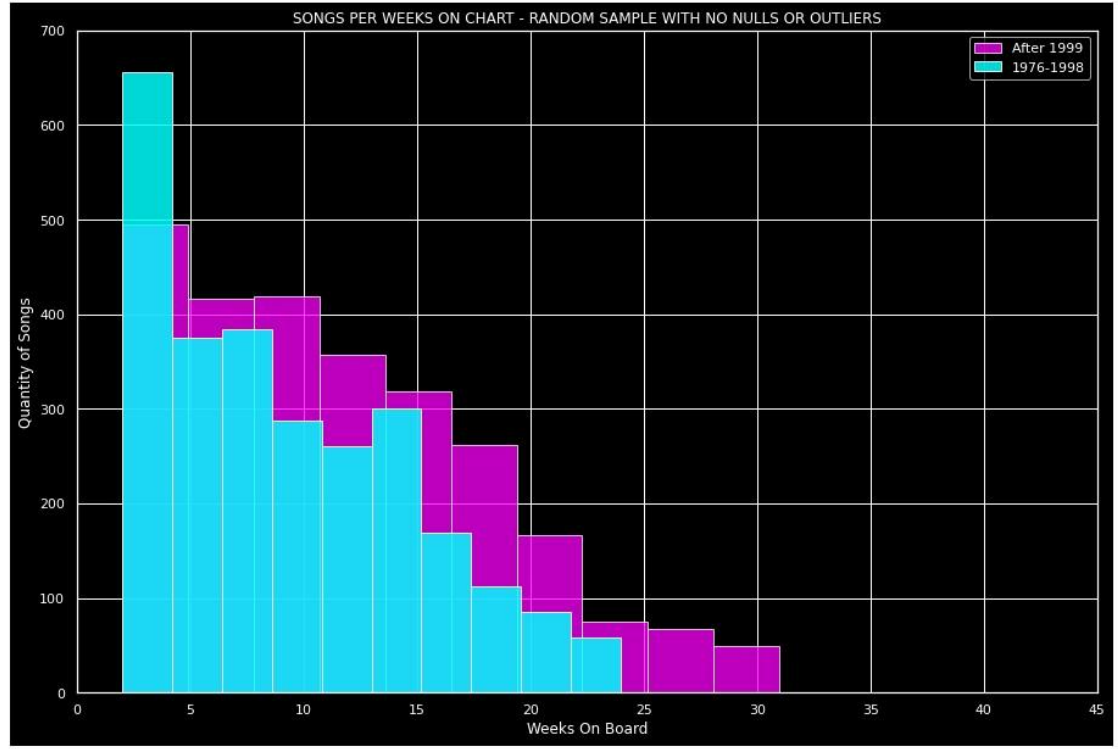
Pre & Post: Shapiro-Wilk

Shapiro-Wilk tests on cleaned data indicated the variable distributions as significantly *different than normal*



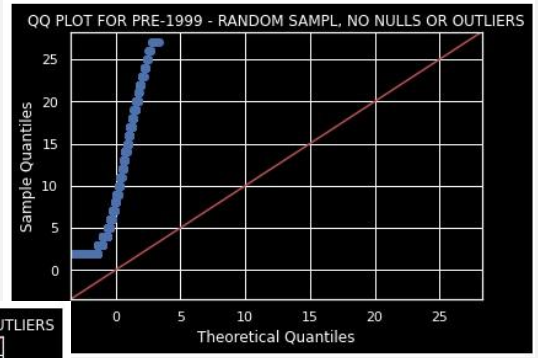
The Color and Shape of: The Data

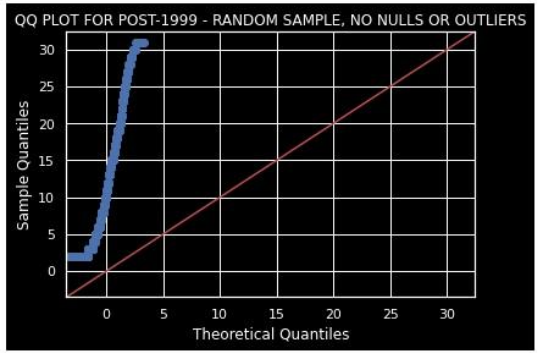
Comparing songs/weeks-on-board for both Post-1999 and Pre-1999 also indicates **NON-NORMAL distribution** (not a bell curve)

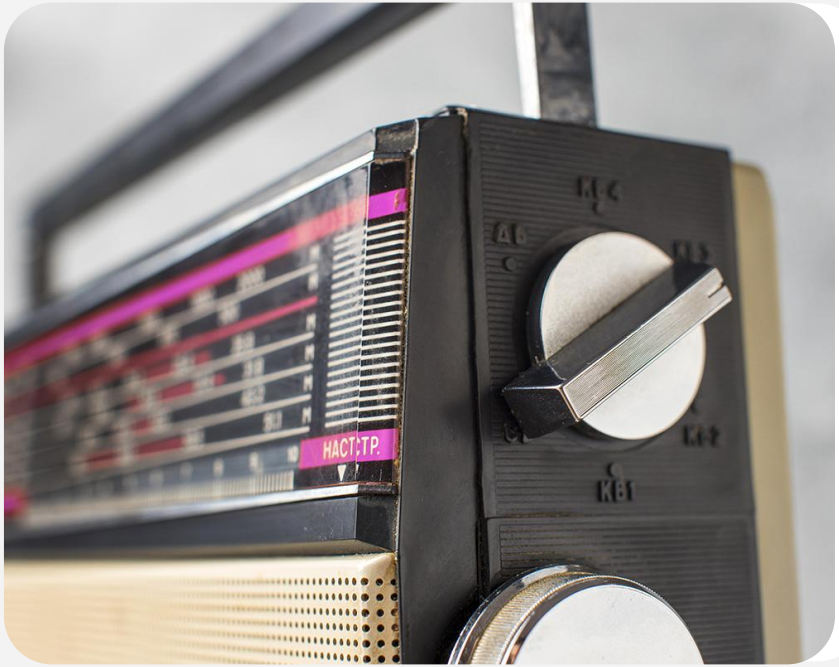


The Color and Shape of: The Data

QQ Plots also indicate **NON-NORMAL distribution** (doesn't follow diagonal line)







KEEP DATA WEIRD

Analysis

Data isn’t weird, it’s just not-normal, sometimes.

Kruskal-Wallis Test

(non-parametric testing for non-normal distributions)

Based on statistical analysis, including the Kruskal-Wallis test, it can be concluded that...

Yes, songs released after 1999 DID chart longer than songs released between 1976 and 1998.

Extensive testing did indicate that the null hypothesis: "H_0 - Songs charted from 1976-1998 did not have significantly different chart times than songs released after 1999" should be rejected.

Rockin’ the Sure Shot

How confident are these assumptions?

Calculating at a 95% confidence interval, it can be determined that:

Post-1999 songs stayed on the charts on average about 2 - 2.5 weeks longer than Pre-1999 songs. *

* This has been determined through many, many random sample data iterations and an averaging of upper and lower confidence limits per sample run.

Recommendations

CONCLUSION: It's clear that the onset of digital file sharing and streaming services in 1999 affected how long songs chart.

Stakeholders should consider how they can optimize the advantages of digital music content delivery to make data-informed decisions regarding content distribution, marketing and revenue generation, and audience development.

Stakeholders

Record Labels

Artists Bands

Streaming Music Services

Traditional Broadcast Radio

Satellite Radio

Why Digital ? ... the advantages to optimize

- **Equitable access (mobile, on-demand, infrastructure)**

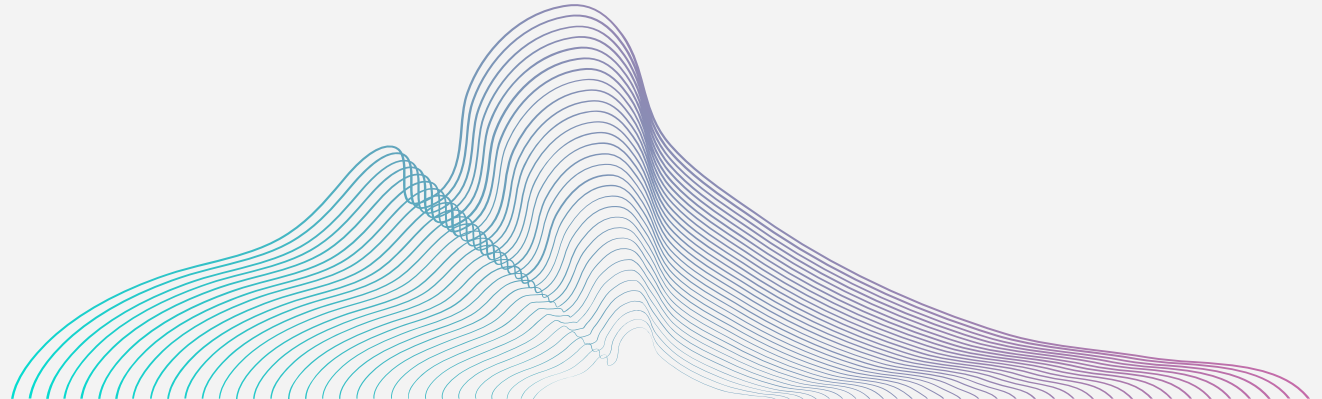
- **Quantifiable exposure and engagement metrics**

Three Areas To Explore

Audience Development

Marketing & Revenue Generation

Content Distribution



Audience Development

What role does a long-charted song and digital content delivery play in a digital age?

The goal of audience development is to build a relationship with listeners, gain new listeners, and develop loyal listeners.

Lengthy-charted songs can help ensure that listeners find joy, comfort, and familiarity in the music and keep coming back for more.

- Radio clock programming - builds loyal listeners ⇁
- Advertising/subscriptions - builds marketable audience + ⇁
- Equitable access - multi-platform access means larger audiences ⇁
- Quantifiable metrics - more-precise audience and advertising reach/ROI measurement

Marketing & Revenue Generation

What role does a long-charted song and digital content delivery play in a digital age?

Familiar long-charted songs keep listeners engaged and returning, which, in turn, **builds a marketable audience for advertisers.**

Live music - **generates more ticket sales** due to more exposure and engagement.

Contrarily, **music sales** have been negatively affected from digital content delivery.

Do not recommend investing in music sales production or marketing.

Revenue % for RIAA	2002	2021
Physical Music Media	98%	11%
Digital Music Media	0%	87%

riaa.com/u-s-sales-database

Content Distribution

What role does a long-charted song and digital content delivery play in a digital age?

Artists/Bands - help determine

- audience/fan preference
- new music trends in style/genre
- if an artist/band should consider changing their ‘sound’ (for example - go “disco” like The Doors and Rolling Stones, or “electronic” like U2 and Madonna)

Album/song releases - help record labels and artists determine **when to release new tracks**

wired.com/2017/05/secret-hit-making-power-spotify-playlist

Recommended Further Research

Implement predictive modelling to determine if there are patterns or correlations between previous lengths of time on the charts and future potential hit songs. Can current long-charting songs predict an estimated time-on-chart of future hits?

Examine what happens around a breakpoint of weeks 20-22 for both pre-1999 and post-1999. Is there something in the data that can help define this, or is this an external influence (legacy industry trend)?

Explore the music styles of pre-1999 and post-1999 to see if genres (especially new genres) influenced time-on-charts. Did songs/artists with more of a hip-hop sound, or more of a electronic sound, produce longer-charting hits? Was there a pattern or correlation in the style of music that play a role in time-on-charts?

Thanks!

Lisa Girard
lisaraygirard@gmail.com

Full data analysis and commentary at:
https://colab.research.google.com/drive/1sjB5VYui3sY2tprQepIQ_VLBuB0gYSOK?usp=sharing


slidesgo