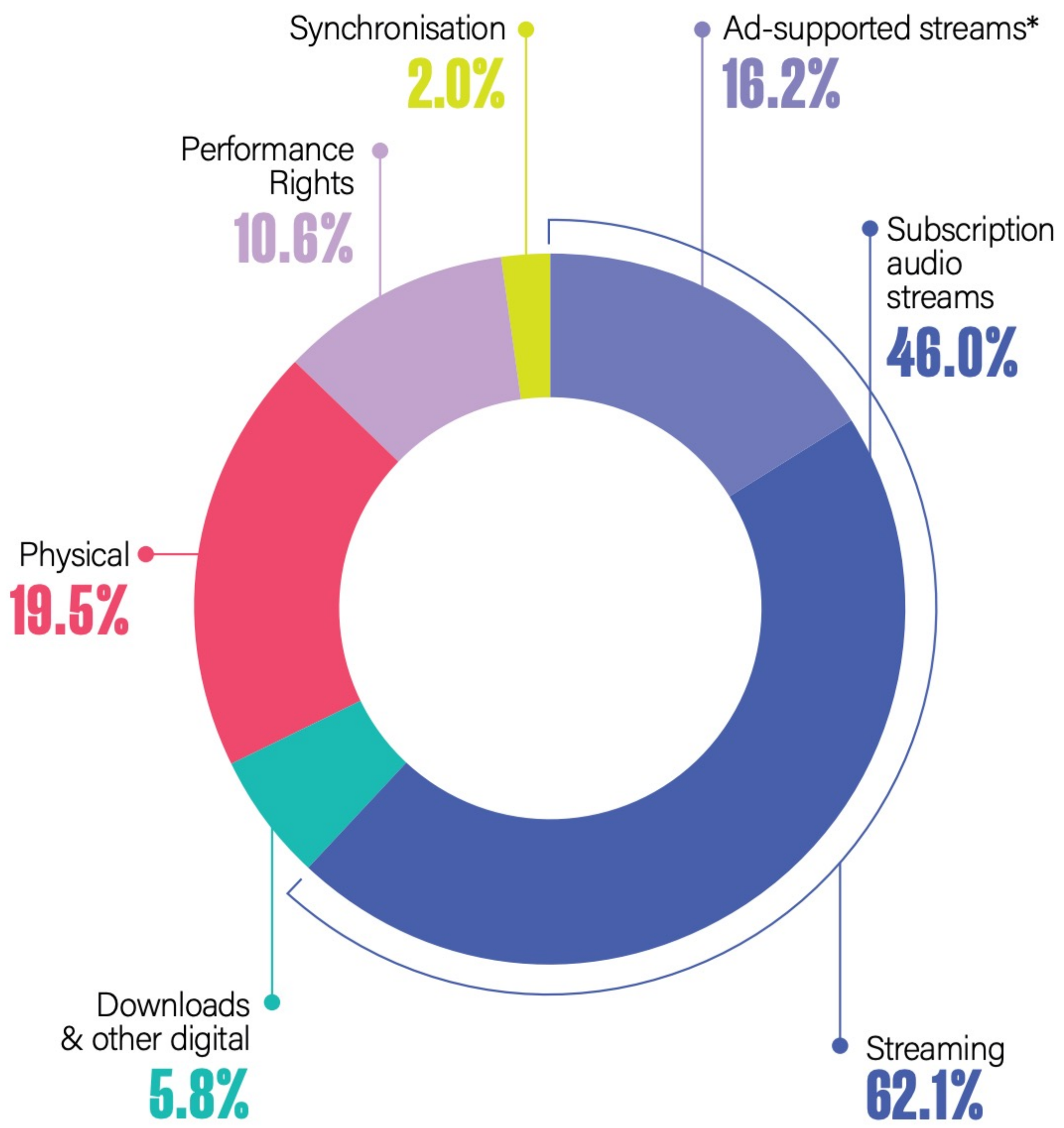
Do Audio Features Have A Causal Effect on Chart Placement?

Causal Inference, Fall ‘21

Louis A. Gomez, Ayesha Parveen

GLOBAL RECORDED MUSIC REVENUES BY SEGMENT 2020



Streaming platforms are the most popular source of music consumption by audiences.

This change also influences how the popularity of a song is calculated billboard

HOT 100

	SONG	ARTIST
1	**My Universe**	Coldplay X BTS
2	**Stay**	The Kid LAROI & Justin Bieber
3	**Industry Baby**	Lil Nas X & Jack Harlow
4	**Way 2 Sexy**	Drake ft. Future & Young Thug
5	**Fancy Like**	Walker Hayes
6	**Bad Habits**	Ed Sheeran
7	**good 4 u**	Olivia Rodrigo
8	**Kiss Me More**	Doja Cat ft. SZA
9	**Knife Talk**	Drake ft. 21 Savage & Project Pat
10	**Levitating**	Dua Lipa

chart dated October 9, 2021

The Billboard Hot 100 charts keep track of the music consumption of songs weekly

Streaming has the highest weight in calculating this ranking

Research Question

In this work, we investigate the causal effects of audio features on chart placement?

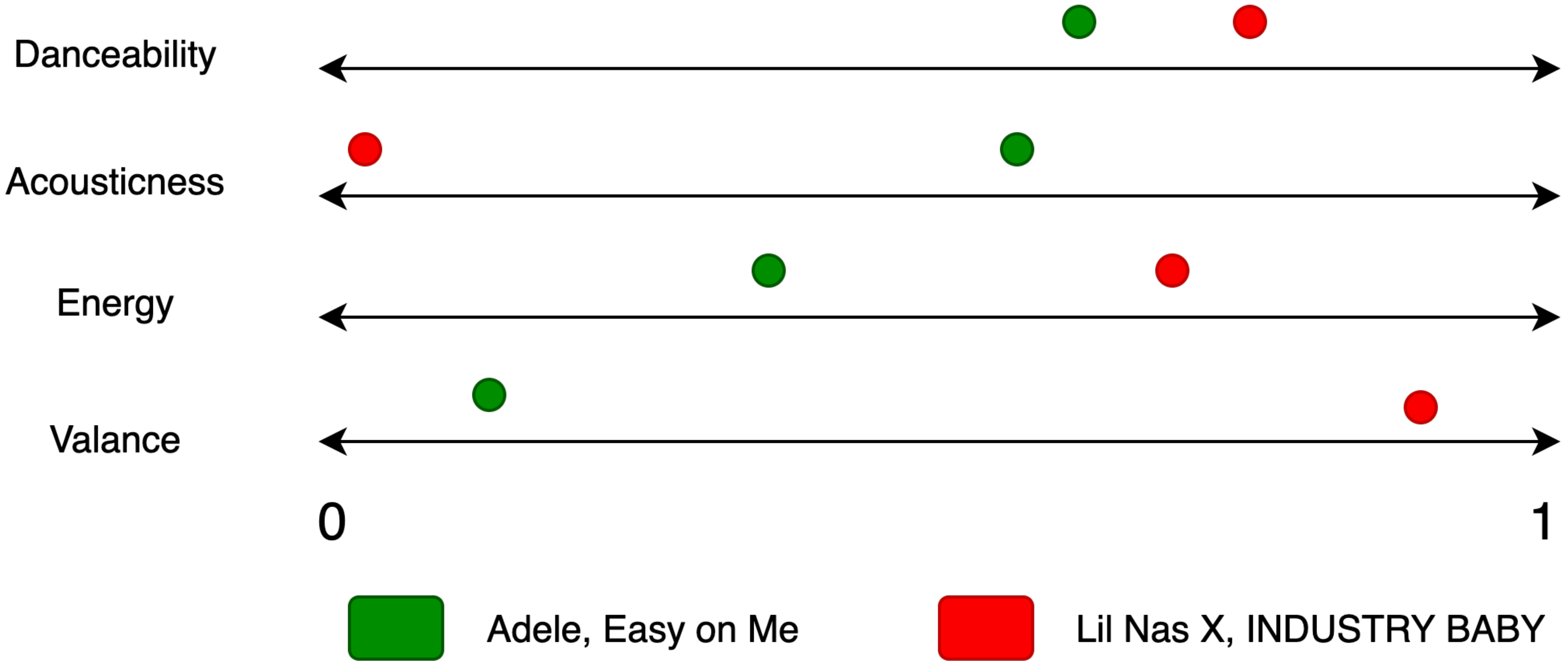
Methods : Data Collection - Billboard

	Postion on chart	Titles	Artists	Weeks	lastPosition	chart release date
0	1	Humble.	Kendrick Lamar	3	3	2017-05-01
1	2	Shape Of You	Ed Sheeran	15	1	2017-05-01
2	3	That's What I Like	Bruno Mars	14	2	2017-05-01
3	4	DNA.	Kendrick Lamar	1	0	2017-05-01
4	5	Mask Off	Future	9	7	2017-05-01

We collected past 5 years (2021-2017) of data from the Billboard charts for the songs that were released in the summer

We only considered songs that had debut for the first time in the charts.

Methods: Data Collection - Spotify



Other audio features are: instrumentalness, mode, liveness, tempo, duration, speechiness and genre

Audio Features	Description
acousticness	A confidence measure to track whether the is acoustic or not.
danceability	Suitability of the track for dancing based on tempo, rhythm stability, beat strength, and overall regularity and activity.
energy	Represents a perceptual measure of intensity and activity.
instrumentalness	Predicts whether a track contains no vocals
mode	Indicates modality (major or minor) of a track
liveliness	Detects the presence of an audience in the recording
valence	The musical positiveness conveyed by a track
tempo	The overall estimated tempo of a track in beats per minute
duration	The duration of a track in milliseconds
speechiness	Detects the presence of spoken words in a track

Methods: Data Processing

After data collection, we have 1197 songs in our dataset

Numerical Features: Discretize using recommended thresholds

Categorical Features (Genre): Map to high level and one hot encode

Outcomes variable (chart position): 1 if > 50 else 0

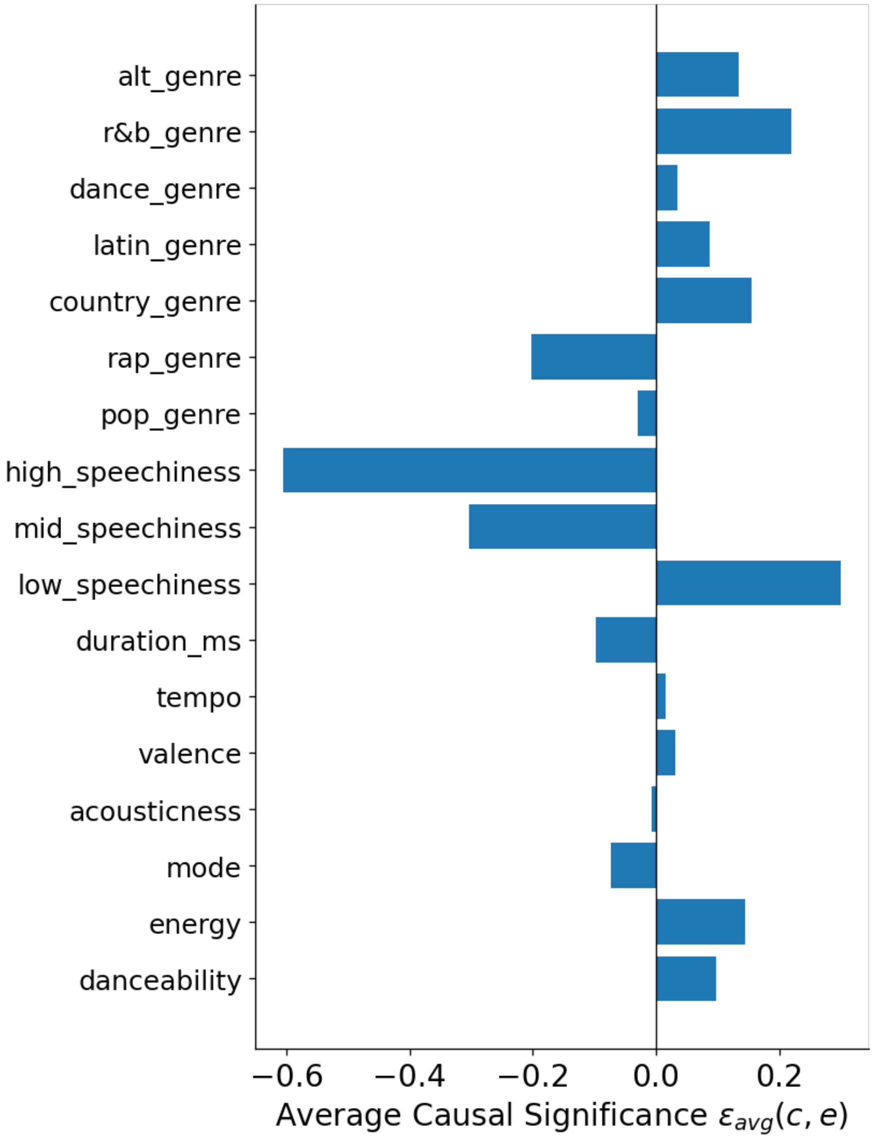
Methods: Causal Inference

Goal: Investigate the causal effects of audio features on chart placement using 17 audio features

$$\mathcal{E}_{avg}(c,e) = \frac{\sum_{x \in X \setminus c} \mathcal{E}_x(c,e)}{|X \setminus c|}$$

$$\mathcal{E}_x(c,e) = P(e|c \wedge x) - P(e|\neg c \wedge x)$$

Results



Estimated using data from 2017 - 2020

Additionally, to evaluate causal relationships, we perform a classification task using features that are non-zero and achieved AUROC of 0.66

Train on (2017 – 2020), test on 2021 data

Limitations

Feature variety – artist popularity

Data collection and data processing – mismatch in naming conventions across databases

Positivity problem

Future Work

Tracking performance after debut – what are the effects after debut

Inclusion of other metrics – Tik Tok performance, artist popularity

Causal Effects in different seasons – winter vs summer

Questions?