BrainStation®

Demo Day

Over the past few months, I've been working on my Data Science Capstone Project: An Exploratory Data Analysis on how audio features play a role on popularity scores of specific artists and genres.

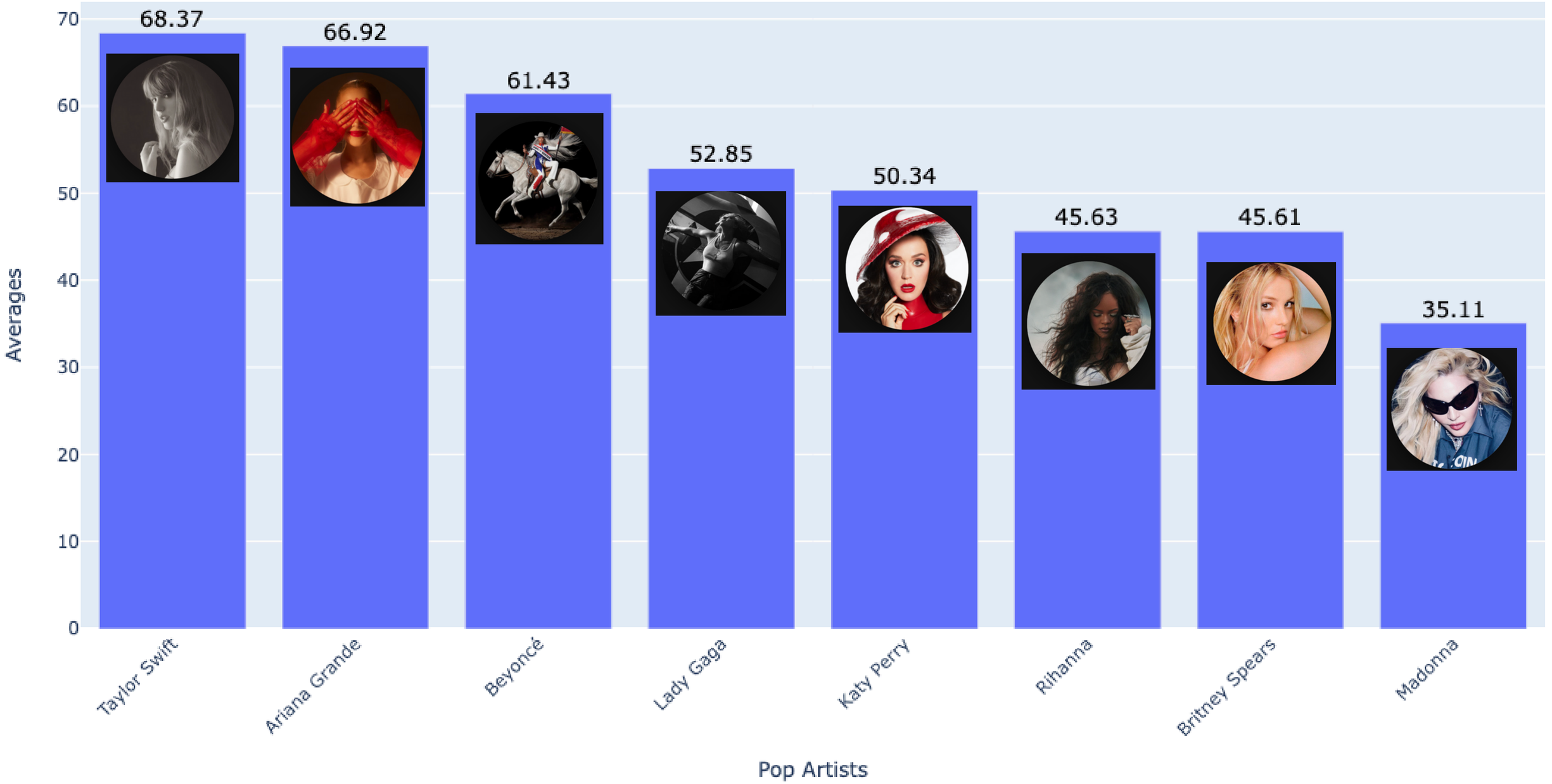
In a nutshell, I created an algorithm that tells you if you can predict the popularity of a new song of a popular artist, based solely on statistical methods that include the following audio features: danceability, tempo, “happiness”, acoustic and more, and using web-scraped data from studio albums available on Spotify, I included the following six genres: Pop, Reggaeton, Electronic, Rap, Classical, and Rock.

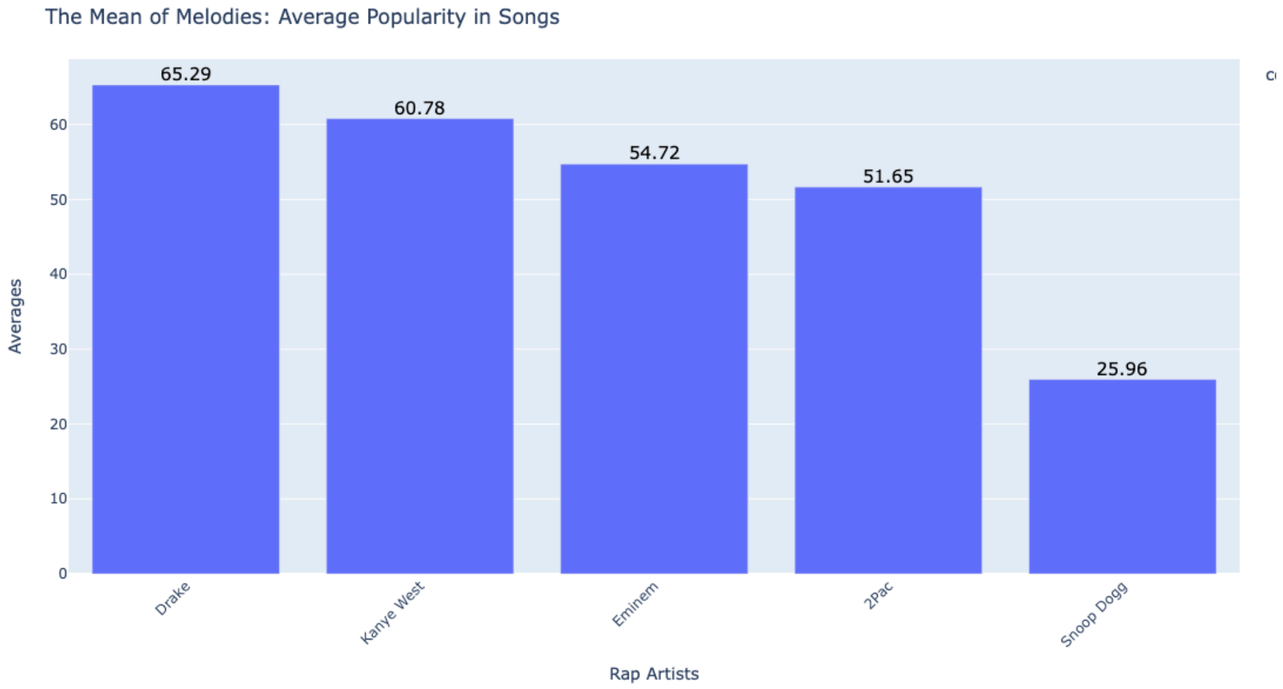
Data Snippet

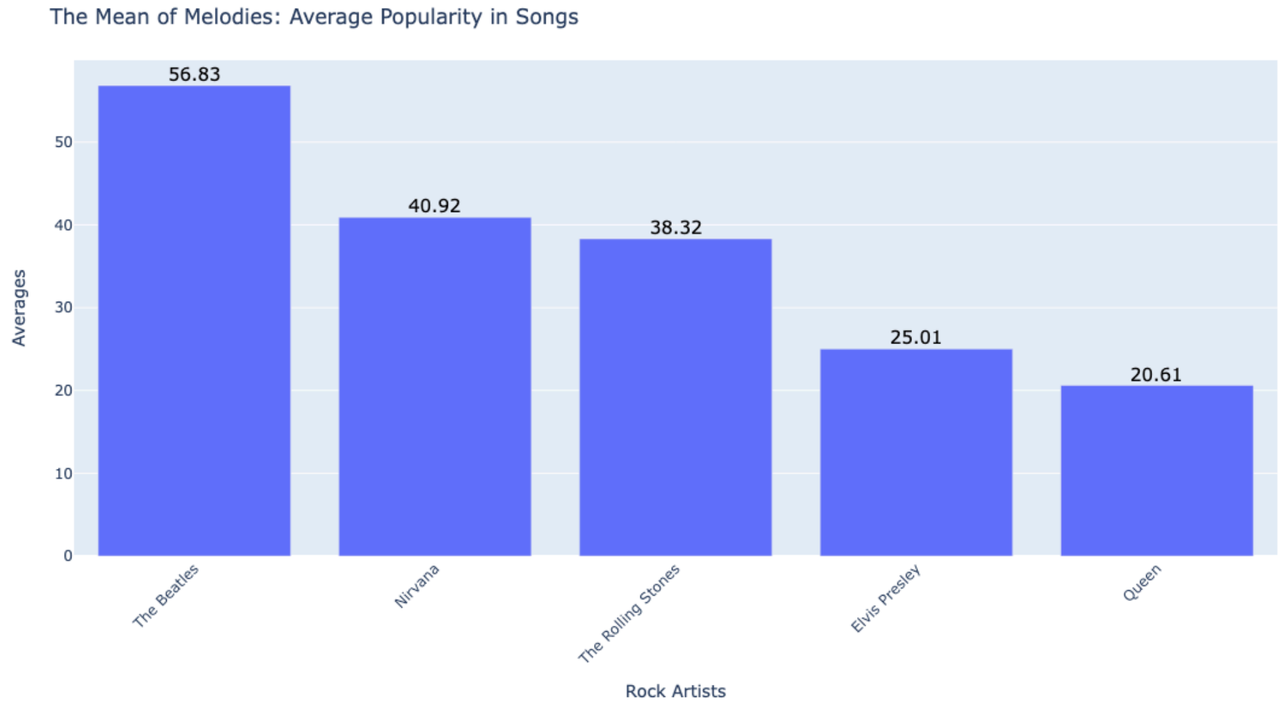
My column of popularity is the target I want to predict

	Name	Artists	Album	popularity	danceability	energy	loudness	acousticness	instrumentalness	Genre
0	Cruel Summer	Taylor Swift	Lover	96	0.552	0.702	-5.707	0.117	2.06	Pop
1	we can't be friends	Ariana Grande	eternal sunshine	95	0.645	0.663	-8.305	0.0657	3.18	Pop
2	Oops!...I Did It Again	Britney Spears	Oops!... I Did It Again	81	0.751	0.834	-5.444	0.3	1.77	Pop
3	We Found Love	Rihanna	Talk That Talk	85	0.734	0.766	-4.485	0.025	0.00138	Pop
4	Run the World (Girls)	Beyoncé	4	72	0.732	0.899	-4.237	0.00496	4.64	Pop

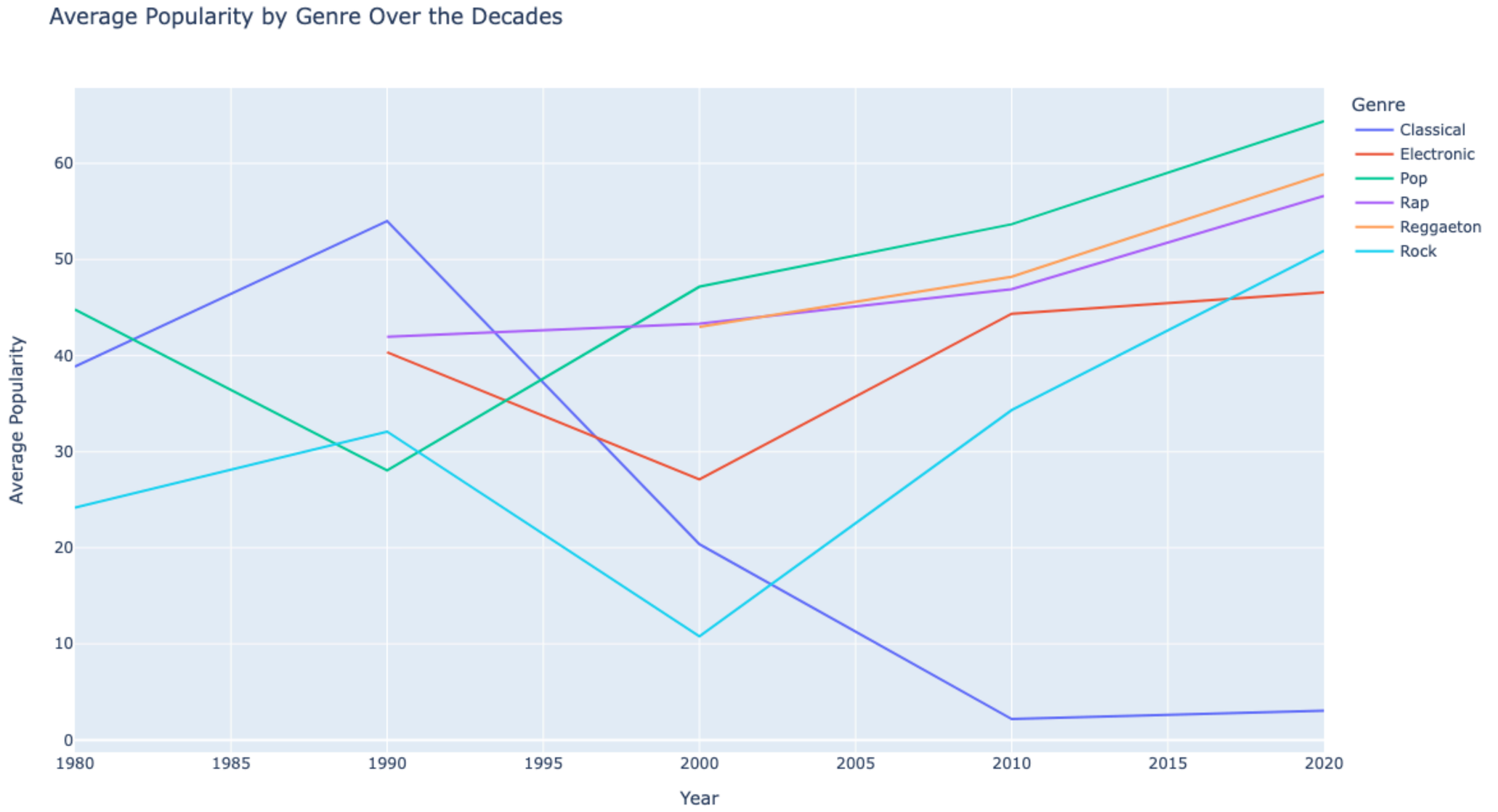
The Mean of Melodies: Average Popularity in Songs




The Mean of Melodies: Average Popularity in Songs
65.29
60.78
54.72
51.65
25.96
Averages
0
10
20
30
40
50
60
Drake
Kanye West
Eminem
2Pac
Snoop Dogg
Rap Artists

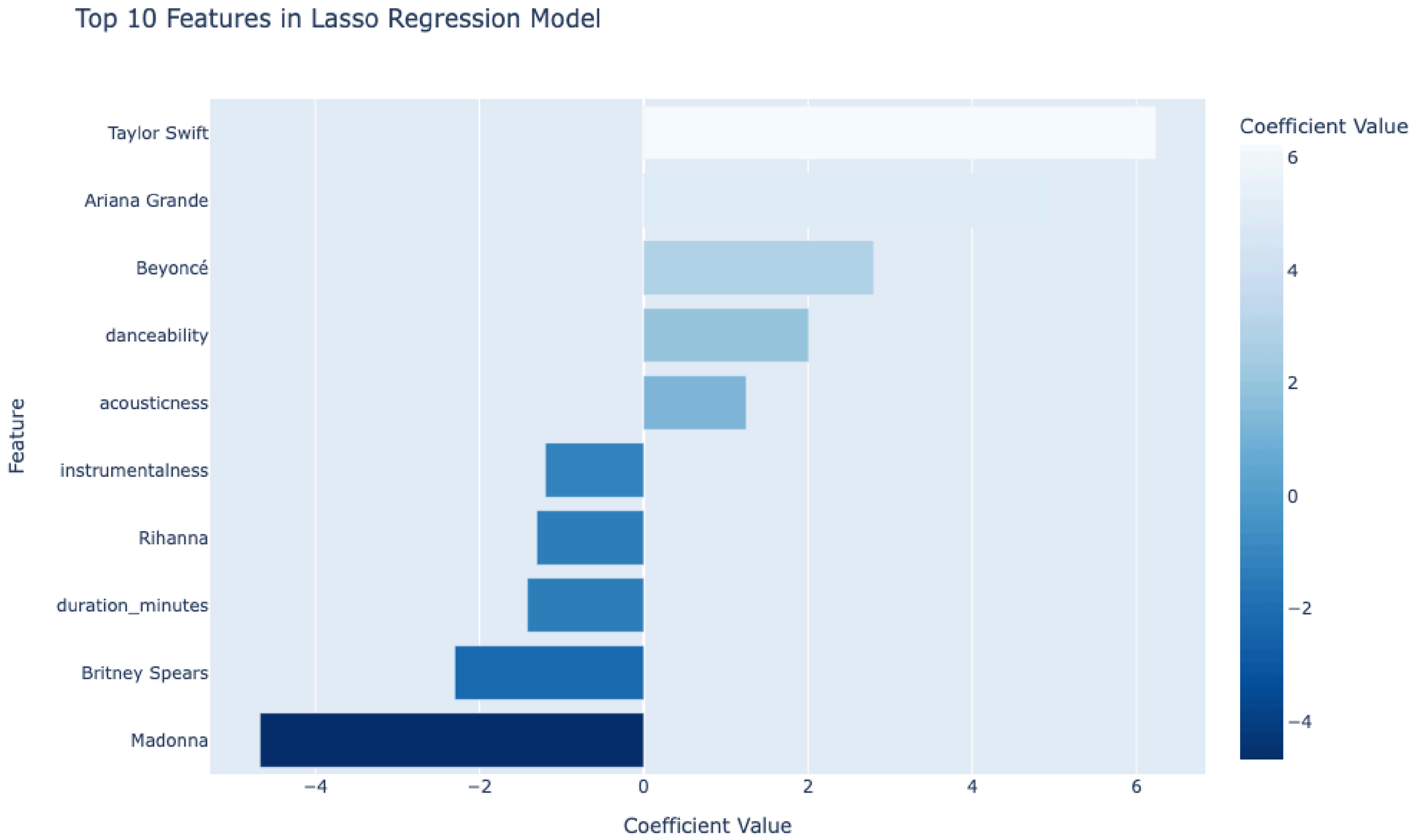

The Mean of Melodies: Average Popularity in Songs
56.83
40.92
38.32
25.01
20.61
Averages
0
10
20
30
40
50
The Beatles
Nirvana
The Rolling Stones
Elvis Presley
Queen
Rock Artists

People listened more symphonies made before 2000.
People didn't listened to Rock on albums dated between and 1995 and 2005.
Reggaeton as the newest genre.



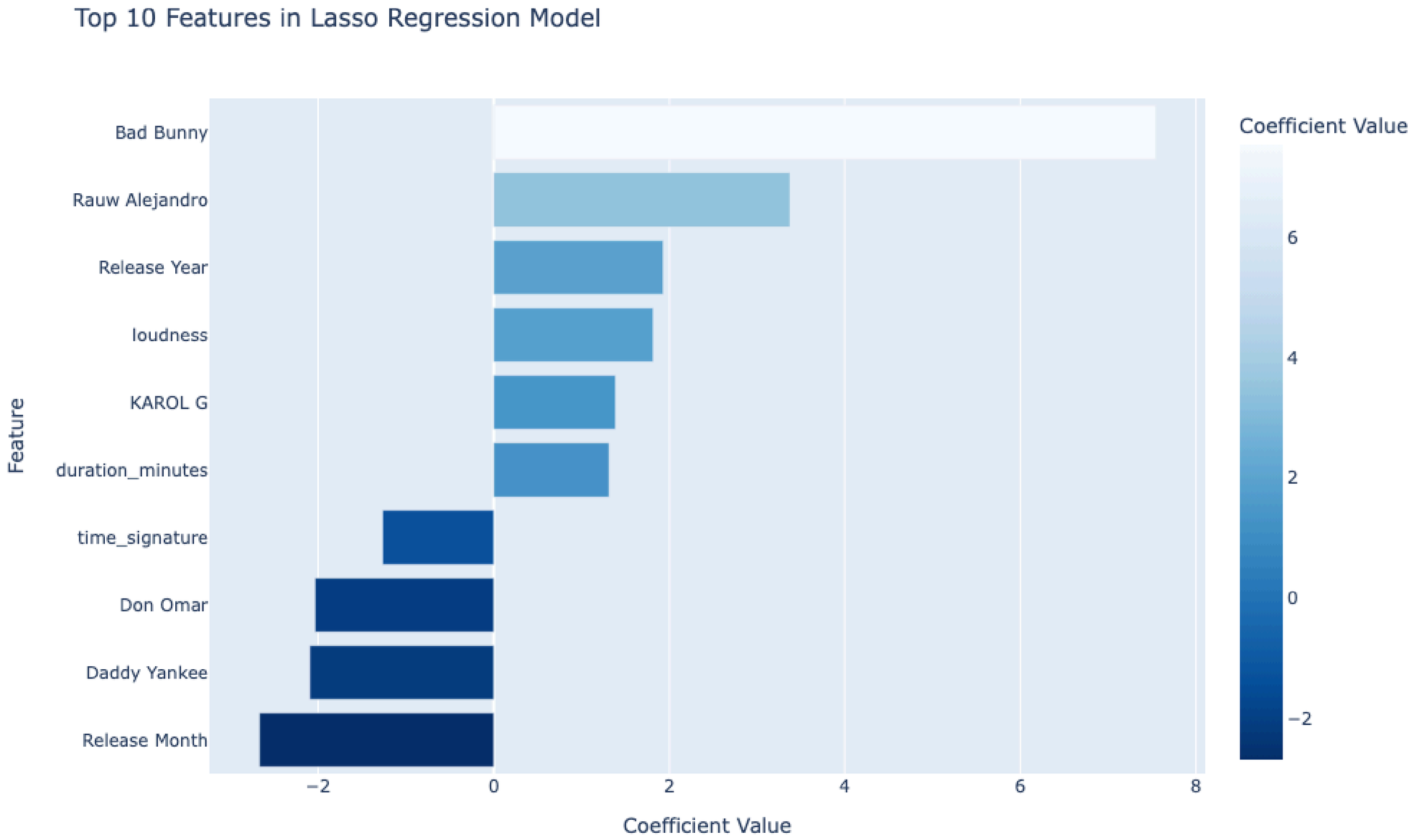
Pop top 5 coefficients and bottom 5 coefficients
Postive and negative correlations

Test Score: 0.590147260697486
MSE Score: 136.55492973958164



Test Score: 0.4853902975288159
MSE Score: 148.06760042032178

Reggaeton top 6 coefficients and bottom 4
Positive and negative correlations



Prediction Scores
(Popularity)

	Pop	Reggaeton	Electronic	Rap	Classical	Rock	All 6 Genres
Lasso (Statiscal Method)	0.58	0.48	0.43	0.65	0.76	0.53	0.25
Ridge (Statiscal Method)	0.59	0.47	0.43	0.65	0.76	0.53	0.25

Top and bottom coefficient per genre, in relation with Audio Features.

Based on the scores of this data sample!	Pop	Reggaeton	Electronic	Rap	Classical	Rock
Do	Danceability	Loudness	Energy	Loudness	Danceability	Duration
Do not do	Instrumentalness	Release Month	Instrumentalness	Acousticness	Month	Instrumentalness

Top and bottom coefficient per artist, in relation with the artists on the dataset.

(I converted the artist into numerical value, so the machine
can include them as well on the regressions).

Based on the scores of this data sample!	Pop	Reggaeton	Electronic	Rap	Classical	Rock
Do	Taylor Swift	Bad Bunny	Daft Punk	Drake	Vivaldi	The Beatles
Do not do	Rihanna	Don Omar	David Guetta	Snoop Dogg	Beethoven	Queen

Discoveries

Derived from scores computed through advanced statistical methodologies applied to the Spotify dataset, discerning insights await both data analysts and stakeholders within the music industry.

1. For optimum prognostication of a track's popularity, meticulous genre-based organization of datasets proves higher scores 🙂 E.G. Do not mix genres to model.

2. Enhanced accuracy in data extraction is contingent upon methodical organization.🙂 E.G. if is studio album or live album, clean the data.

3. For practitioners in the music creation, a prudent strategy involves consistent album releases. Regardless of vocal powers, the relevance of an artists is sustained through periodic updates.🙂 E.G. Artists releasing more albums lately proved more popularity of those with a strengthens in their voice.

4. Furthermore, for aspiring artists, adherence to the audio features that are you top coefficient, is key for success.🙂

5. In a further analysis I would like to only used studio albums from the past 3 years, web scrape new artists and old artists, and web scrape data from other Apps like YouTube Music and Apple. 🙂

BrainStation®

Thank you!

Please feel free to reach out at my
linkedin.com/in/victorcornejoo

Extra slide I did not use.

Top and Bottom Coefficients

Based on the scores of this data sample!	Pop	Reggaeton	Electronic	Rap	Classical	Rock
Do	1.Taylor Swift 2.Ariana 3.Beyonce 4.Danceability 5.Acousticness	1. Bad Bunny 2.Rauw Alejandro 3.Release Year 4.Loudness 5.Karol G 6.Duration	1.Daft Punk 2.Energy 3.Avicii 4.Loudness 5.Release Year	1.Drake 2.Kanye West 3.Loudness 4.Month 5.Day 6.Eminem 7.Danceability	1.Vivaldi 2.Danceability 3.Energy 4.Key	1.The Beatles 2.Nirvana 3.Duration 4.Loudness
Do not do	1.Instrumentalness 2.Songs with long duration 3.Rihanna 4. Britney 5. Madonna	1.Don Omar 2.Daddy Yankee 3.Release Month	1.David Guetta 2.Instrumentalness 3.Tiesto 4.Speechiness 5.Duration	1.Snoop Dogg 2.Acousticness 3.Day	1.Month 2.Beethoven 3.Day 4.Loudness 5.Year 6.Valence	1.Queen 2.Year 3.Elvis 4.Day 5.Instrumentalness 6.Liveness