Discovering the Next Super Bowl Headliner

Final presentation

Group3:

Xiaoyi Huang

Alanna F Correa

Wendy Ralston

Travis D Nieves

Brandon F Vasquez

Agenda

1

Introduction & Domain Background
Xiaoyi Huang

2

Abstraction
Alanna Correa

3

Methodology & Tools
Wendy Ralston

4

Insights
Travis Nieves

5

Reflection & Conclusion
Brandon Vasquez

Xiaoyi Huang

Introduction & Domain Background

Introduction

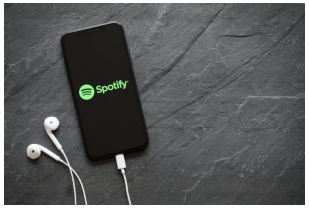
Using a data-driven approach, our music-based marketing agency is crafting a pitch to identify the artist with the most significant potential to captivate the massive Super Bowl audience

Dataset Foundation

title	artist	top genre	year	bpm	nrgy	dnce	val	dur	spch	pop
Hey, Soul Sister	Train	rock	2010	97	89	67	80	217	4	83
Love The Way Y	Eminem	hip hop	2010	87	93	75	64	263	23	82
TiK ToK	Kesha	dance pop	2010	120	84	76	71	200	14	80
Bad Romance	Lady Gaga	dance pop	2010	119	92	70	71	295	4	79
Just the Way Yo	Bruno Mars	pop	2010	109	84	64	43	221	4	78
Baby	Justin Biebe	pop	2010	65	86	73	54	214	14	77
Dynamite	Taio Cruz	dance pop	2010	120	78	75	82	203	9	77
Secrets	OneRepubli	dance pop	2010	148	76	52	38	225	4	77
Empire State of	Alicia Keys	hip hop	2010	93	37	48	14	216	3	76

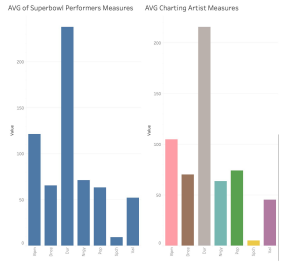
Based on Spotify's top streamed songs (2010-2019), we analyze track data including artist, popularity, genre, and audio features to inform our strategy

Consumption Evolution



Recognizing the shift from album purchases to streaming, our data-driven analysis identifies relevant artists based on actual listening behavior quantitatively

Visualizing the Pitch



Bridging the gap between data science and marketing, we use Python and Tableau to make complex music data accessible and actionable for any marketer

Alanna Correa

Abstraction

Data Domain Background

Billboard Ranking

- Data Sources: Combines sales, streaming, and radio airplay
- Ranking Criteria: Based on song performance across platforms
- Key Data Variable: Billboard ID (unique song identifier)

Data Organization:

- Organized by Ranking and Year

Spotify Insights

- Song Data: Title, Artist, Genre, Year, Duration
- Audio Features: BPM (Beats per minute), DB (Loudness in decibels)
- Ranked Features: Energy, Danceability, Liveliness, Valence, Acousticness, Speechiness, Popularity (not in order)

Column1	title	artist	top genre	year	bpm	nrgy	dnce	dB	live	val	dur	acous	spch	pop
1	Hey, Soul Sister	Train	neo mellow	2010	97	89	67	-4	8	80	217	19	4	83
2	Love The Way You Lie	Eminem	detroit hip hop	2010	87	93	75	-5	52	64	263	24	23	82
3	TiK ToK	Kesha	dance pop	2010	120	84	76	-3	29	71	200	10	14	80
4	Bad Romance	Lady Gaga	dance pop	2010	119	92	70	-4	8	71	295	0	4	79
5	Just the Way You Are	Bruno Mars	pop	2010	109	84	64	-5	9	43	221	2	4	78
6	Baby	Justin Bieber	canadian pop	2010	65	86	73	-5	11	54	214	4	14	77
7	Dynamite	Taio Cruz	dance pop	2010	120	78	75	-4	4	82	203	0	9	77
8	Secrets	OneRepublic	dance pop	2010	148	76	52	-6	12	38	225	7	4	77
9	Empire State of Mind (Part II) Broken Down	Alicia Keys	hip pop	2010	93	37	48	-8	12	14	216	74	3	76

Data Abstraction

Data Types

Structurally working with a file containing Rows (**Items**) and Columns (**Attributes**)

Dataset

Original dataset is preserved in a CSV format; translating to a **Table** of **Items** and **Attributes**

Attributes

Data can be organized by Categorical attributes of "Title," "Artist," and "Top Genre and Quantitative attributes of BPM, Energy, Danceability, Valence, Duration, Speechiness, and Popularity

	title	artist	top genre	ye	bpm	nrgy	dnce	val
1	Hey, Soul Sister	Train	rock	2010	97	89	67	80
2	Love The Way Yi	Eminem	hip hop	2010	87	93	75	64
3	TiK ToK	Kesha	dance pop	2010	120	84	76	71
4	Bad Romance	Lady Gaga	dance pop	2010	119	92	70	71
5	Just the Way Yo	Bruno Mars	pop	2010	109	84	64	43

Task Domain Background

HOW

Capturing a cultural moment with the resurgence of the '10s music, we identified a data set that will be culturally relevant and assist in narrowing our pitch

WHAT

As a marketing agency, our request for proposal was to identify a select group of artists to propose for the next Super Bowl Halftime show.

WHY

With the shift in streaming based consumption habits; Spotify's dataset would provide us the bridge between current consumer consumption within a given time period of music popularity

Task Abstraction

Foundational Database

Generate a Python encoded notebooks to translate into a Tableau dashboard view that will allow the user ease of access to the data with no coding requirement

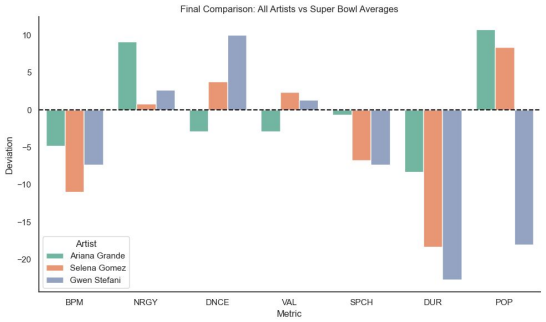
Comparison Benchmarks

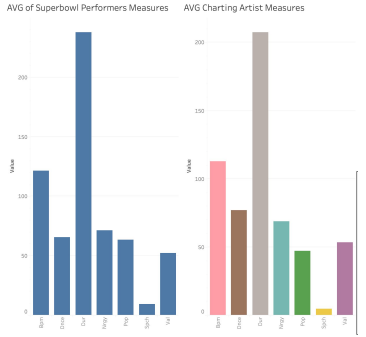
Establish a historical view of artists who have and have not performed at the Super Bowl for context

Audience Interaction

Create filters giving the user dynamic control over story-telling (i.e. did the production company want a pop artist or someone with high danceability?)

	artist	title	bpm	nrgy	dnce	val	dur	spch	pop
0	Bruno Mars	[Just the Way You Are, Marry You, Just the Way...	126.2	74.2	69.7	60.8	224.5	4.9	72.0
1	Coldplay	[Paradise, Atlas - From □The Hunger Games: Cat...	137.2	65.0	42.8	18.0	266.2	3.2	58.8
2	Eminem	[Love The Way You Lie, Walk On Water (feat. Be...	84.5	68.5	61.5	63.0	283.5	23.5	73.5
3	J Balvin	[Mi Gente (feat. Beyoncé)]	105.0	72.0	76.0	47.0	210.0	8.0	68.0
4	Janet Jackson	[BURNITUP!]	123.0	71.0	83.0	83.0	250.0	18.0	31.0
5	Jennifer Lopez	[On The Floor, I'm Into You, Papi, Invading My...	122.0	77.7	69.2	60.6	231.7	13.1	55.9
6	Justin Timberlake	[Suit & Tie, Mirrors - Radio Edit, TKO, Take B...	104.1	65.1	67.7	47.1	297.9	11.9	56.9
7	Katy Perry	[Teenage Dream, California Gurls, E.T., Last F...	132.2	75.9	65.7	59.9	224.2	6.4	62.1
8	Lady Gaga	[Bad Romance, Telephone, Alejandro, Born This ...	120.1	74.9	64.0	49.9	246.6	5.3	68.9
9	Madonna	[Ghosttown, Living For Love]	101.0	67.0	60.5	24.0	234.5	6.5	53.0
10	Maroon 5	[Misery, Moves Like Jagger - Studio Recording ...	118.4	69.3	70.0	57.1	217.6	4.9	74.9
11	Missy Elliott	[WTF (Where They From), Pep Rally]	105.0	78.5	88.0	44.0	228.0	16.0	52.5

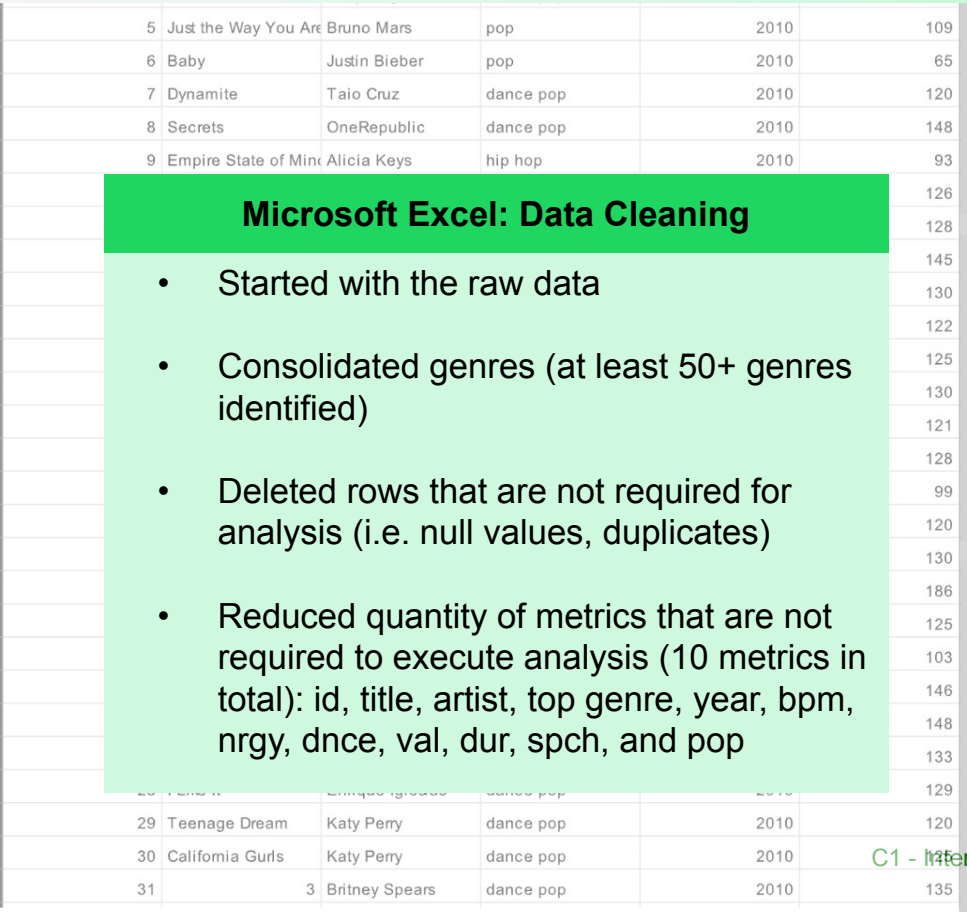




Wendy Ralston

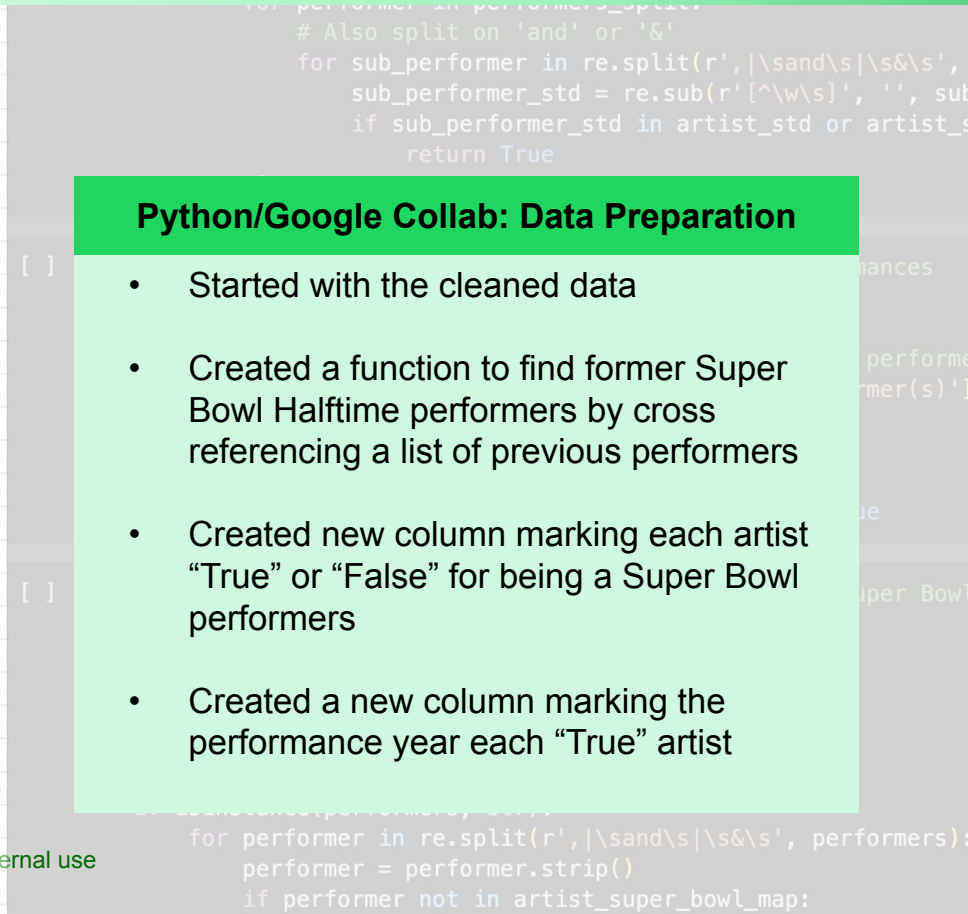
Methodology & Tools

Analysis Tools: Data Cleaning & Preparation



Microsoft Excel: Data Cleaning

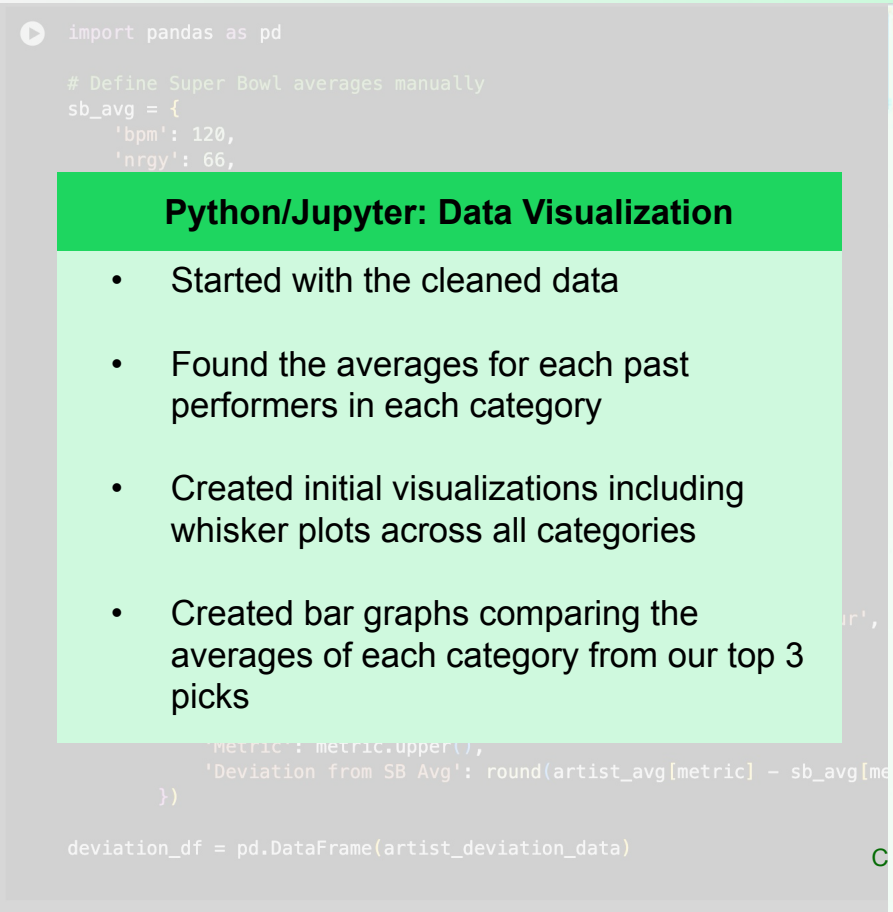
- Started with the raw data
- Consolidated genres (at least 50+ genres identified)
- Deleted rows that are not required for analysis (i.e. null values, duplicates)
- Reduced quantity of metrics that are not required to execute analysis (10 metrics in total): id, title, artist, top genre, year, bpm, nrgy, dnce, val, dur, spch, and pop



Python/Google Collab: Data Preparation

- Started with the cleaned data
- Created a function to find former Super Bowl Halftime performers by cross referencing a list of previous performers
- Created new column marking each artist “True” or “False” for being a Super Bowl performers
- Created a new column marking the performance year each “True” artist

Analysis Tools: Data Visualization



Python/Jupyter: Data Visualization

- Started with the cleaned data
- Found the averages for each past performers in each category
- Created initial visualizations including whisker plots across all categories
- Created bar graphs comparing the averages of each category from our top 3 picks

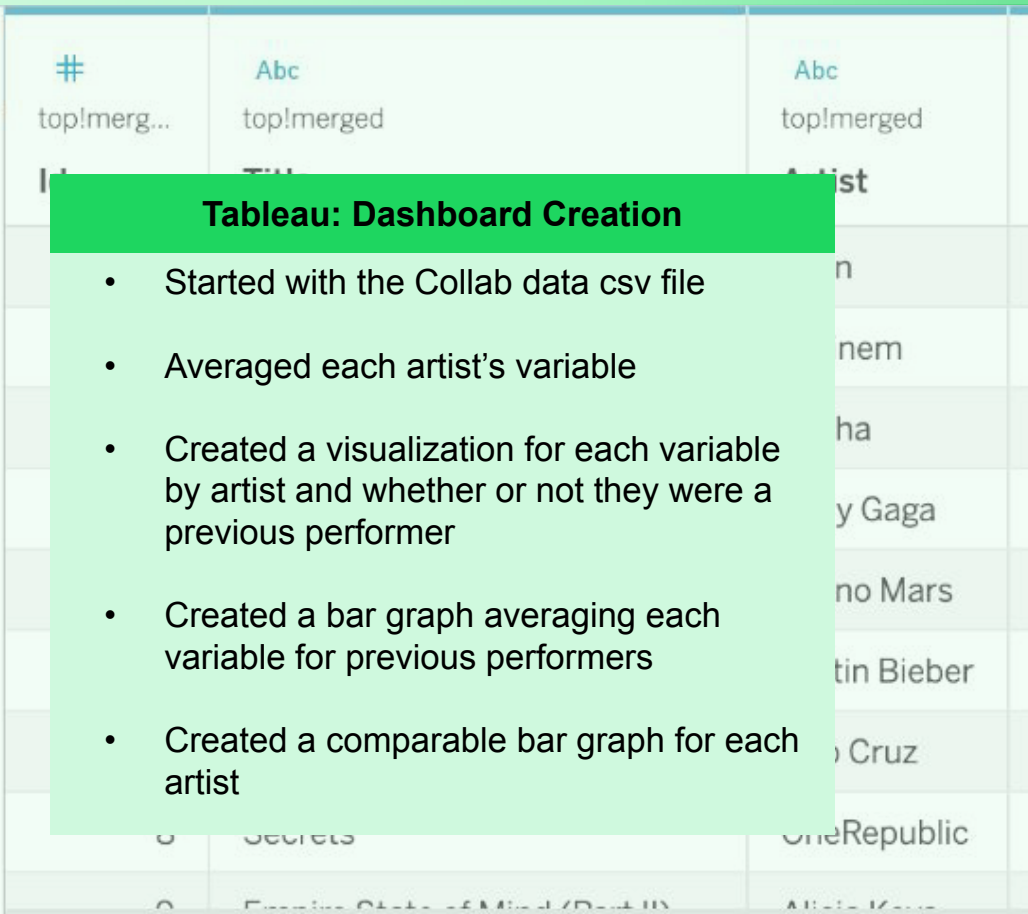


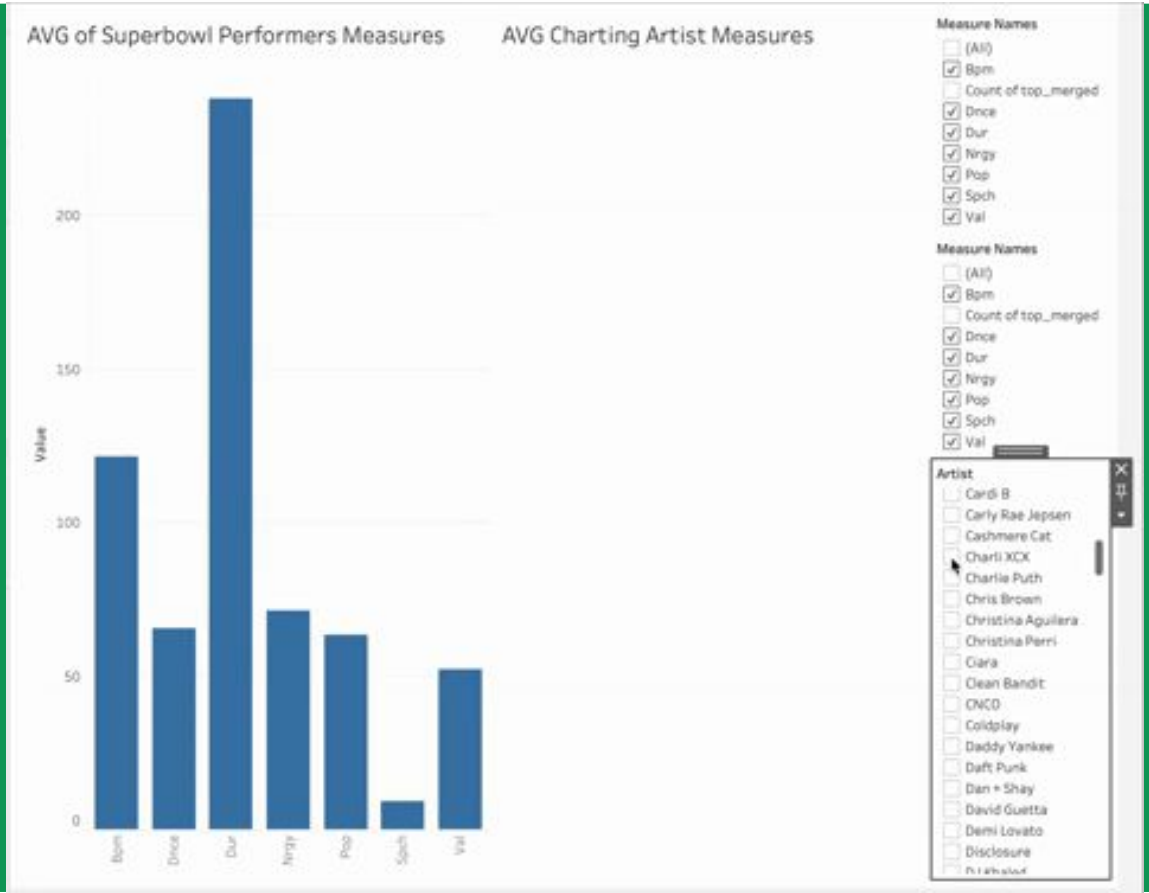
Tableau: Dashboard Creation

- Started with the Collab data csv file
- Averaged each artist's variable
- Created a visualization for each variable by artist and whether or not they were a previous performer
- Created a bar graph averaging each variable for previous performers
- Created a comparable bar graph for each artist

BPM, VAL, POP, and DNC Dashboard

Dur, Spch, and Nrgy Dashboard

Past Performers vs. New Performer Dashboard

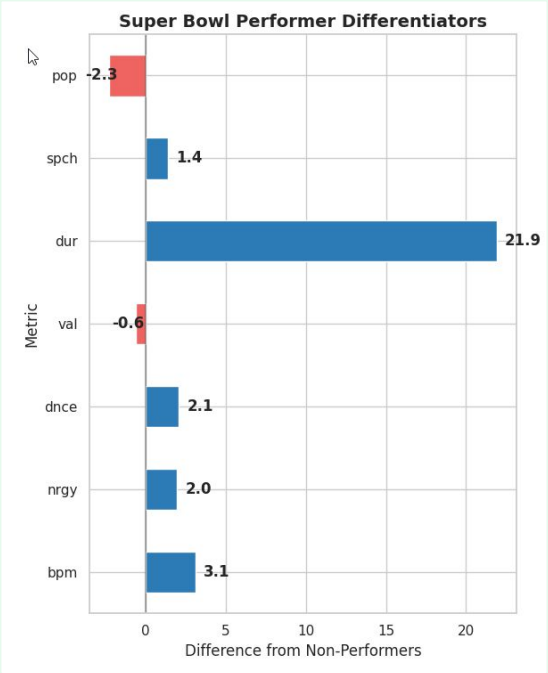


Travis Nieves

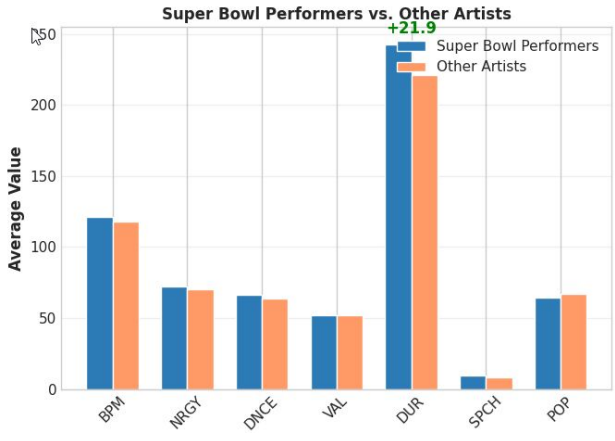
Insights

From Visualization to Insight: What the Data Tells Us

Key Patterns in Performance Data



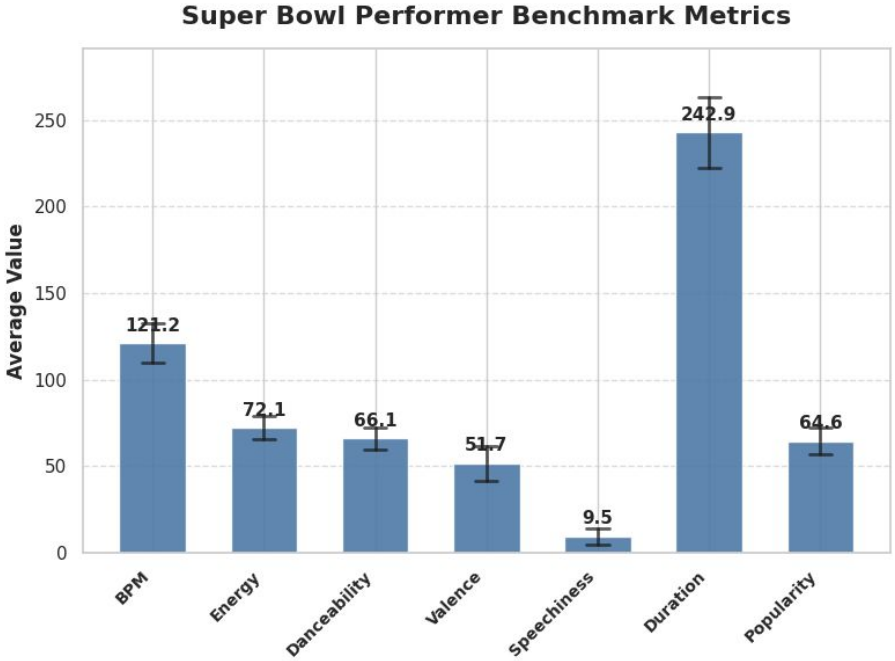
Metric	SB Performers	Others	Difference
BPM	121.2	118.0	+3.1
Energy	72.1	70.2	+2.0
Danceability	66.1	64.0	+2.1
Valence	51.7	52.3	-0.6
Duration	242.9	221.0	+21.9
Speechiness	9.5	8.1	+1.4
Popularity	64.6	66.9	-2.3

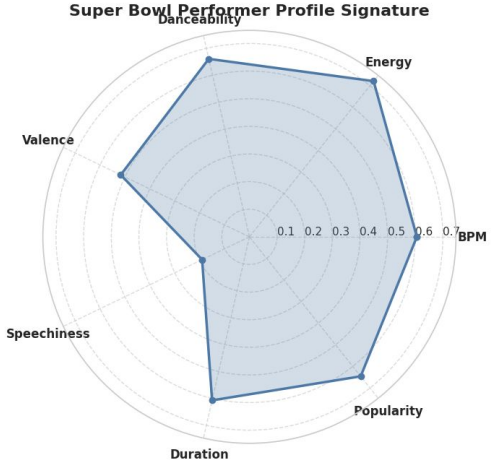


Key Statistical Patterns:

- Super Bowl performers have higher BPM (+3.6), Energy (+2.5), and Danceability (+1.2) compared to non-performers
- They feature longer songs (Duration +26.4 seconds on average)
- They include more vocal elements (Speechiness +2.3)
- Interestingly, they score slightly lower on Popularity (-4.3) and Valence (-1.6)

Creating the Super Bowl Performance Benchmark





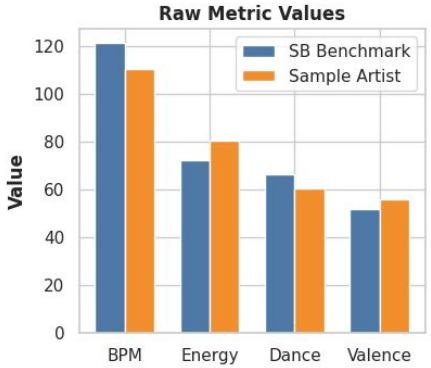
Essential Performance Trio: BPM, Energy, and Danceability show consistent patterns (CV=0.19-0.20) across successful performers, creating the core "Super Bowl sound"

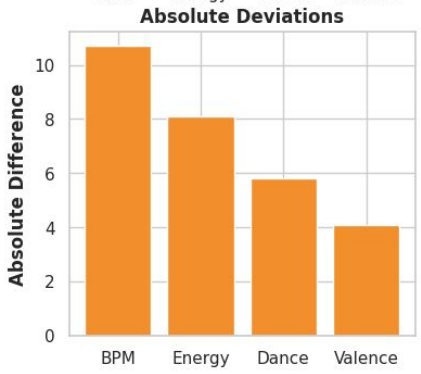
Most Variable Attribute: Speechiness has extremely high variability (CV=0.98), indicating this metric has minimal impact on performance success

Super Bowl Performer Benchmark Values

Metric	Average	Variability	CV
BPM	121.2	±23.2	0.19
Energy	72.1	±13.5	0.19
Danceability	66.1	±12.9	0.20
Valence	51.7	±20.2	0.39
Speechiness	9.5	±9.3	0.98
Duration	242.9	±40.9	0.17
Popularity	64.6	±15.4	0.24

The Deviation Analysis Framework: Measuring Performance Fit



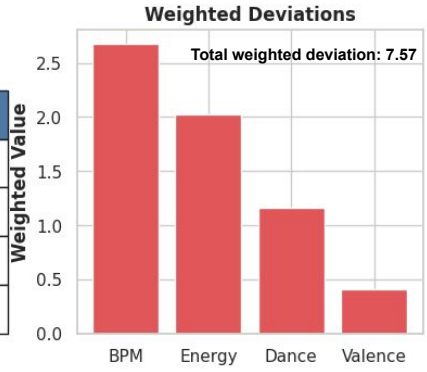


Deviation Calculation

Metric	SB Avg	Artist	Deviation
BPM	121.2	110.5	-10.7
Energy	72.1	80.2	8.1
Dance	66.1	60.3	-5.8
Valence	51.7	55.8	4.1

Metric Weighting System

Metric	Weight
BPM	0.25
Energy	0.25
Dance	0.2
Valence	0.1



Objective Evaluation: Our framework transforms subjective performance qualities into measurable metrics

Weighted Approach: Higher weights assigned to performance-critical factors (BPM, Energy) that drive audience engagement

Benchmark Alignment: Lower total deviation indicates better alignment with successful Super Bowl performances

Comparative Analysis: This methodology allows for fair comparison across different musical styles and eras

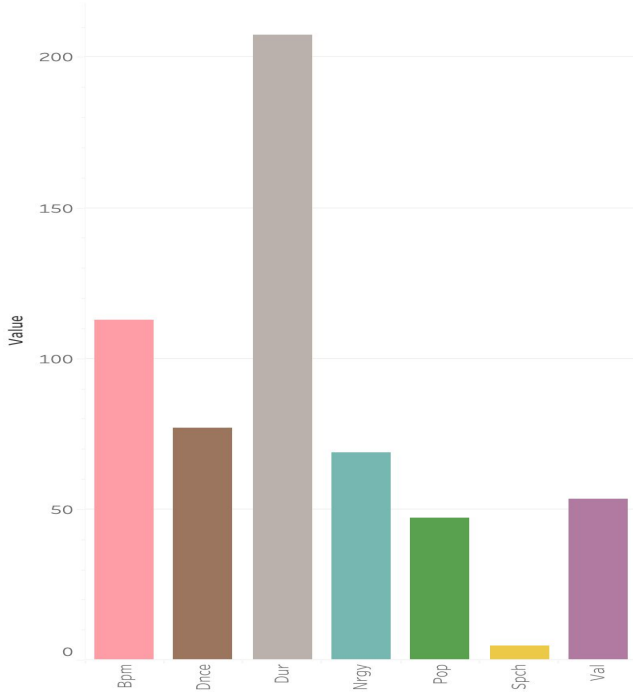
Deviation Interpretation: Negative values show underperformance, positive values show exceeding benchmarks

Brandon Vasquez

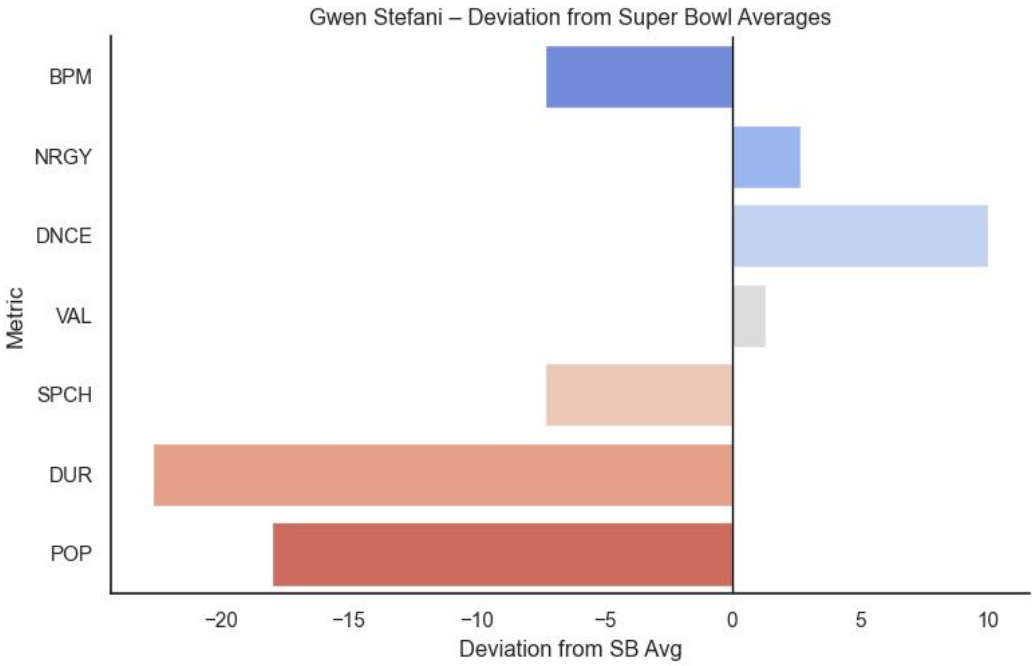
Reflection & Conclusion

Final Artist Recommendations – Gwen Stefani

Artist Metrics: Gwen

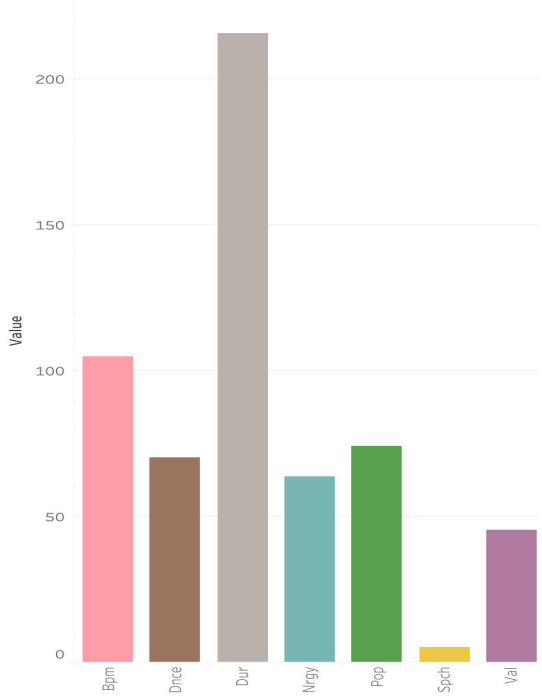


Gwen: Deviation of average metrics

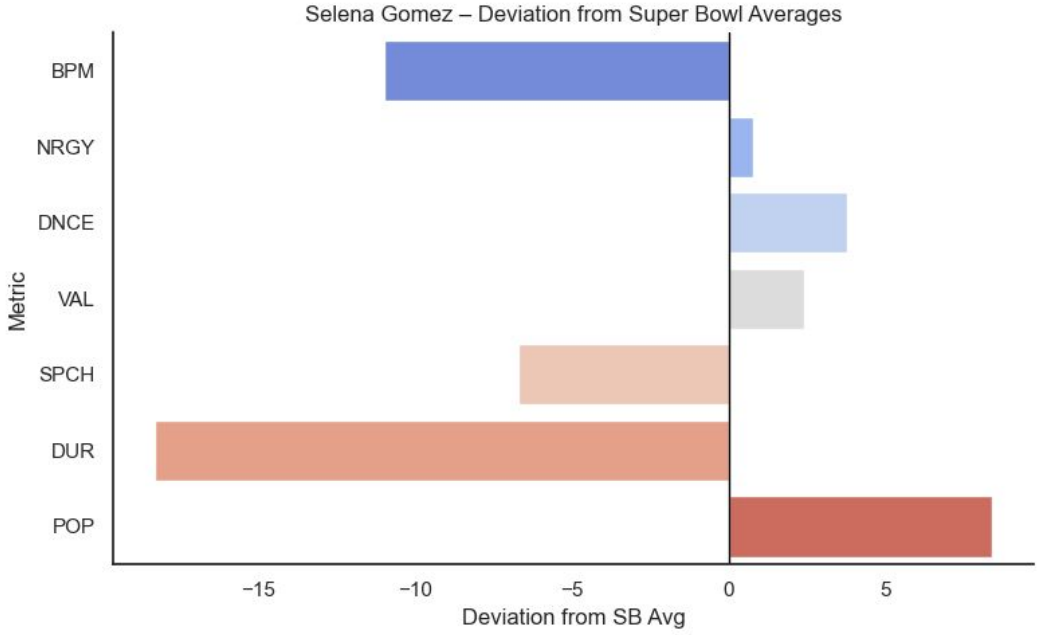


Final Artist Recommendations – Selena Gomez

Artist Metrics: Selena

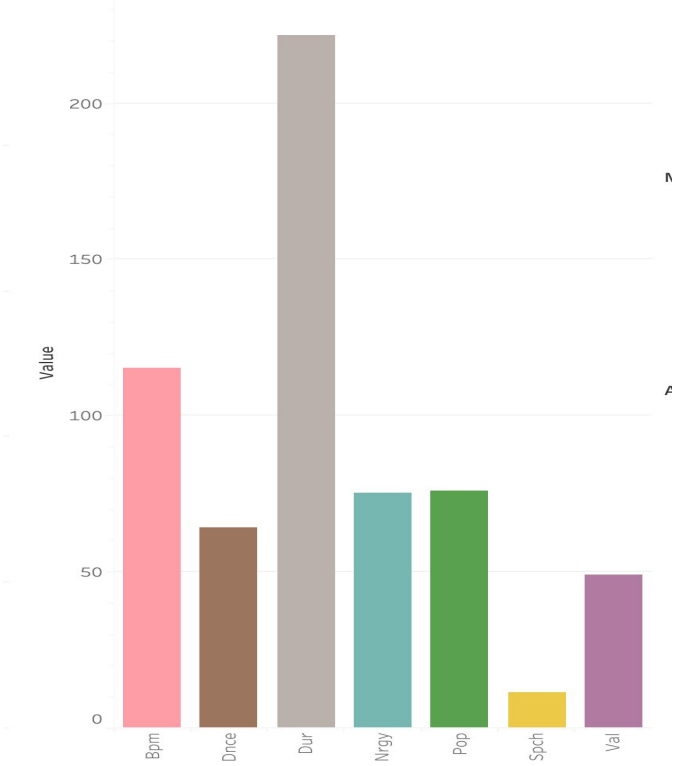


Selena: Deviation of average metrics

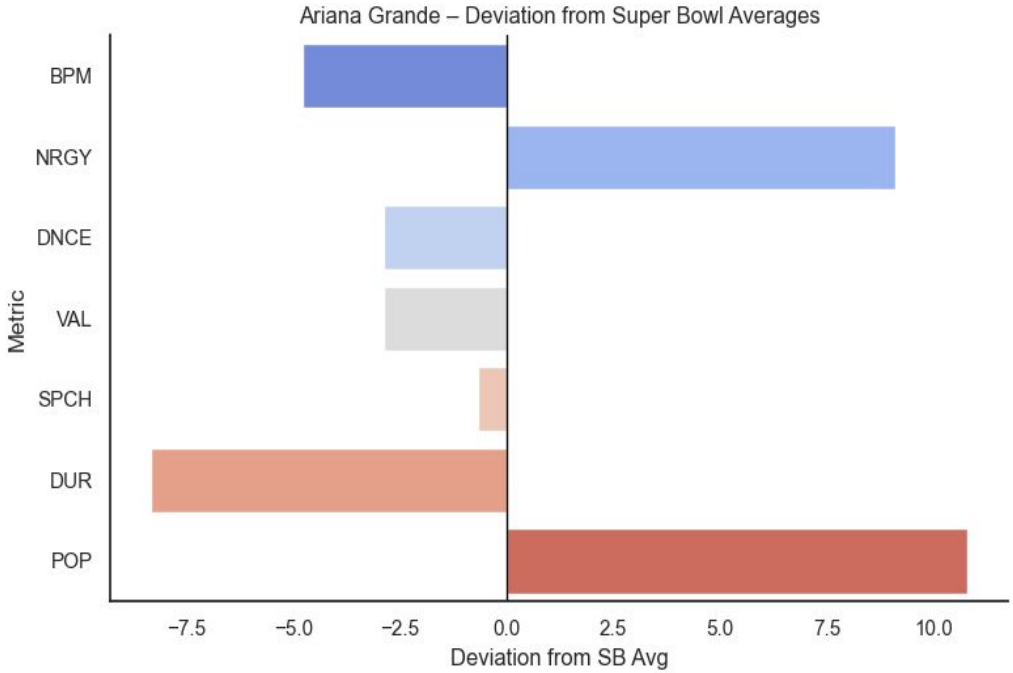


Final Artist Recommendations – Ariana Grande

Artist Metrics: Ariana



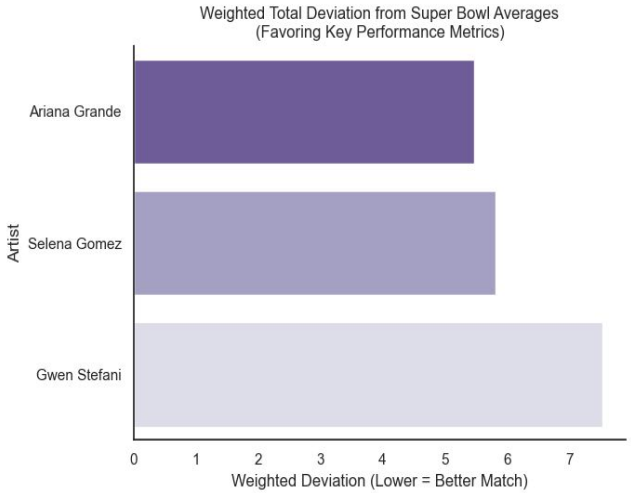
Ariana: Deviation of average metrics



Conclusion

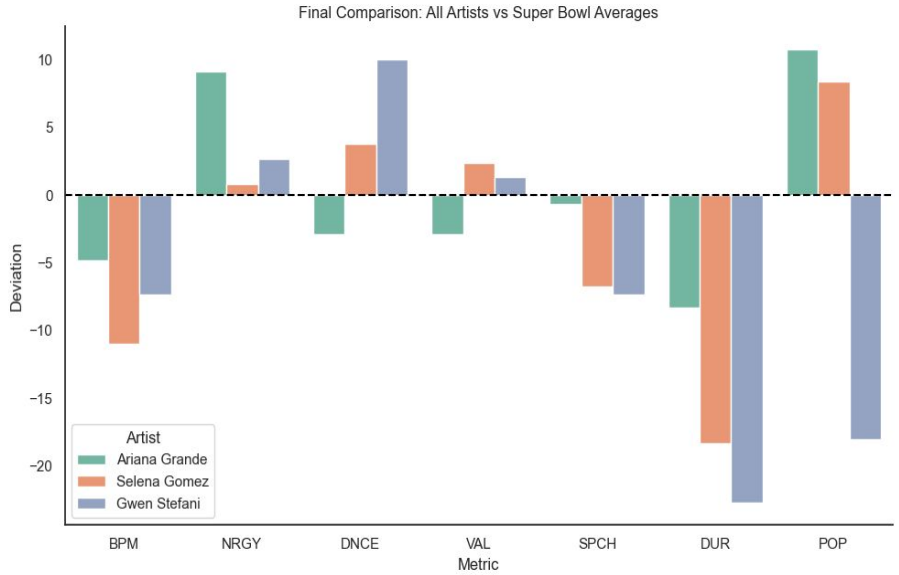
Ariana Grande for Superbowl LX

Benchmark Analysis



Ariana Grande had the lowest cumulative weighted deviation—indicating the strongest alignment with Super Bowl performer standards.

Artist Outperformance



Outperformed Selena Gomez *by 2× in Energy, Danceability, and Valence.*

Outperformed Gwen Stefani *by 3× in BPM, Energy, and Danceability.*

Discussion & Future Work

TAKEAWAY

Strengths

- Took a performance-driven approach—not just popularity or streams
- Used real audio traits that reflect stage impact (energy, danceability, etc.

Weaknesses

- Limited time period for benchmarks skews the historical performance trends
- Didn't factor in logistics like artist booking availability or budget

INSIGHTS

- Not all popular songs are suited for performance (e.g., low-energy chart hits)
- Balanced artists like Ariana thrive when we focus on *how* music feels, not just *how often* it's played

FUTURE WORK

- Extend dataset into the 2020s to reflect new cultural trends
- Incorporate concert footage analysis or live performance ratings
- Add variables like budget, artist scheduling, and tour cycles for real-world viability
- Cross check findings with more tangible dta like spotify and apple music streams, radio streams, google searches etc.

Thank you