

Music Lyrics Data Science



Data Day at the Beach 2025



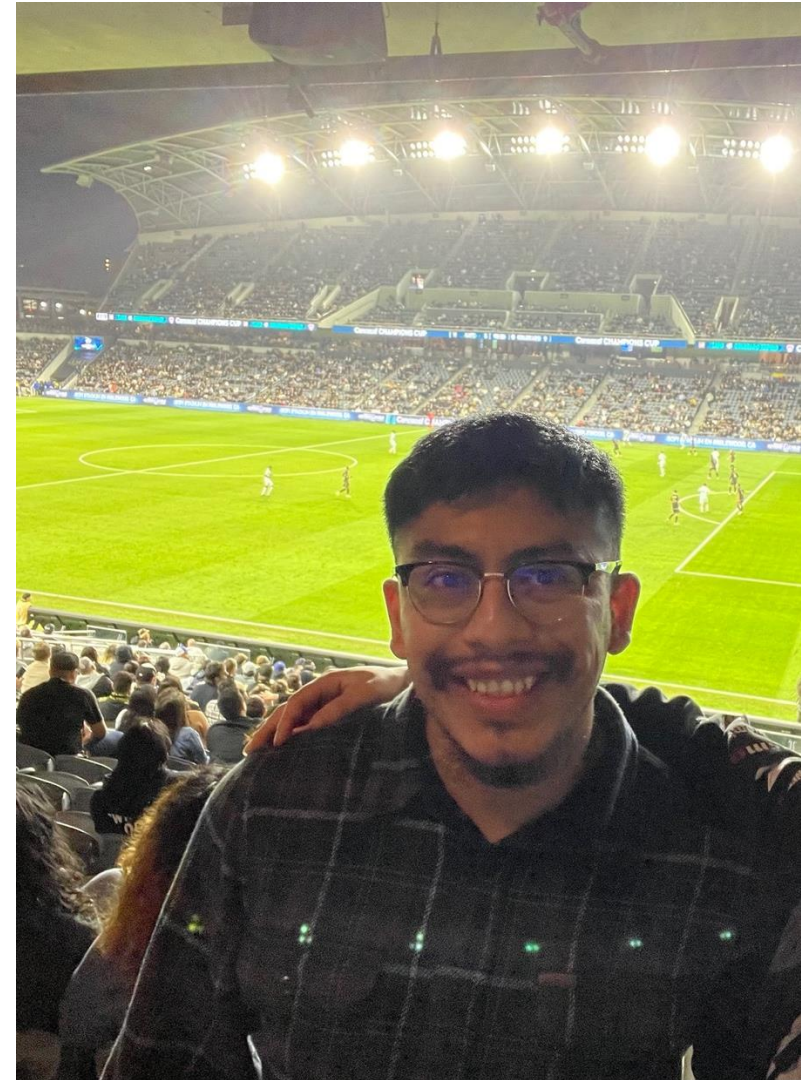
Introduction – Andres Gonzalez

- Background in Statistics, Mathematics and Data Science.
- Experience in R, SQL, Python, and data visualization tools like ggplot2 and Tableau.
- Big fan of soccer, movies, and exploring music across different genres.



Gerry Cruz

- B.S. in Applied Mathematics at CSUN
- Candidate in M.S. Applied Statistics
- Hobbies:Swimming, Latin Dancing, Table Tennis, Gym, Rock Climbing, Comedy
- Coding Experience:Python,R,SQL



Robby Born

- B.S. in Political Science at LMU '19
- Candidate in M.S. Applied Statistics
- Currently: Interning at two non-profits focusing on higher education attainment and public health mapping for evacuation zones
- Coding/Software Experience: Python, R, SQL, SAS, GIS



STAY CONNECTED SLIDE

- Links to more resources to dig deeper (eg any resources in the cloud, github, etc you curate for the participants)
- Links to your LinkedIn
- Coursera/EdX/LinkedIn Learning etc free Courses that could help them go deeper
- Consider creating a linktree (like [this](#)) that will be one link that has all resources in one place
- Your volunteer can help you with these.



OUTLINE/OVERVIEW of the SESSION

- Introduction to the Dataset – Explore the Genius lyrics dataset and understand its JSON structure.
- Data Cleaning & Preprocessing – Remove unwanted text (e.g., [Chorus], [Verse]) and tokenize lyrics into individual words.
- Plot 1: Word Frequency Analysis – Visualize the top 10 most frequently used words from the artist's top 50 songs.
- Sentiment Analysis – Analyze the emotional tone by scoring positive and negative words across songs using the Bing lexicon.
- Plot 2: Sentiment Map of Songs – Create a sentiment map to visualize where each song falls on the positive-negative spectrum.
- Wrap-up & Discussion – Discuss key takeaways and explore how this workflow applies to other types of data.



Billie Eilish Sentiment Analysis

(Positive and Negative)

Unreleased Songs

LL ASLEEP, WHERE DO WE GO?



Visualization by Andres Gonzalez (X @ AndresAnalytics)
Code found at github.com/OKcomputer626

my boy
THE GREATEST
my future
6.18.18
come out and play
What Was I Made For?
8
Six Feet Under
lovely
i love you
Halley's Comet

Negative

Positive

listen before i go
you should see me in a crown
wish you were gay
WILDFLOWER
TV
hostage
THE DINER
Male Fantasy
LUNCH
goodbye
CHIHITO
Billie Bossa Nova
when the party's over
NDA
Therefore I Am
The 30th
BITTERSUITE
SKINNY
idontwannabeyouanymore
everything i wanted
BIRDS OF A FEATHER
watch
COPYCAT
Bored
BLUE
bellyache
party favor
Happier Than Ever
ocean eyes
xanny
Your Power
L'AMOUR DE MA VIE
ilomilo
bury a friend
bitches broken hearts
all the good girls go to hell
No Time To Die
bad guy
my strange addiction

BILLIE EILISH

Sentiment analysis of the 50 most popular songs from Genius,
arranged from the most positive (top left)
to the most negative (bottom right)

