

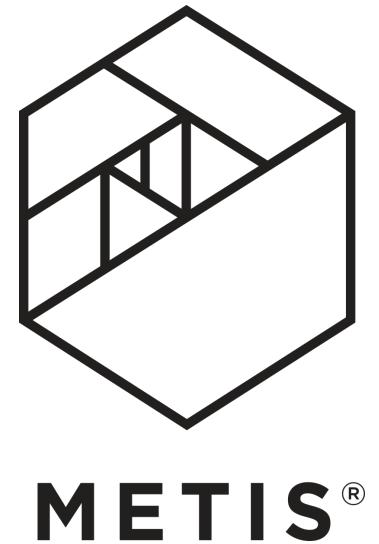
# Argentinian Tango Lyrics

Sentiment & Topics NLP

METIS Data Science and Machine Learning Bootcamp

by Krystian Krystkowiak, 2022

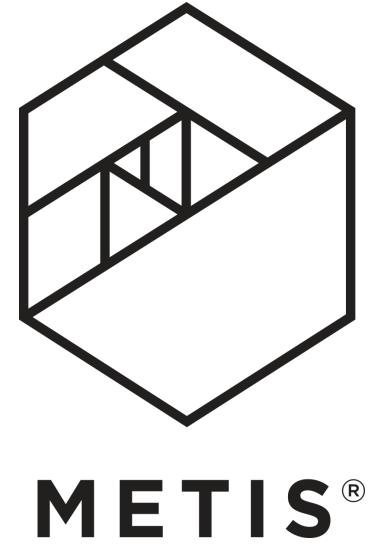
# Introduction



- **Golden Age of Tango Music (1935-1955)**
- 2022 - tango **dance is back**
- But... the **music is old**
- GOAL: **Identify key topics** of tango lyrics **to help modern orchestras** to keep spirit of the period. Insight for **collectors, Dj's, teachers and dancers**



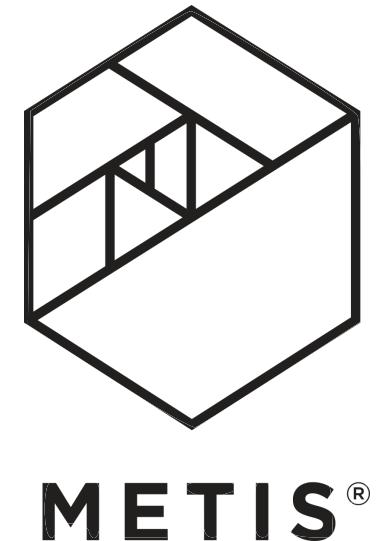
# Methodology



- Dataset of **Golden Age Tango music (1920-1960)** from **el-recoodo.com**, shared on GitHub by user gefero
- Over **2400 songs** in **Spanish** with an average of 145 words per song
- Features include **lyrics** for **NLP analysis**, as well as song **title, author, orchestra, singer and year**



# Methodology

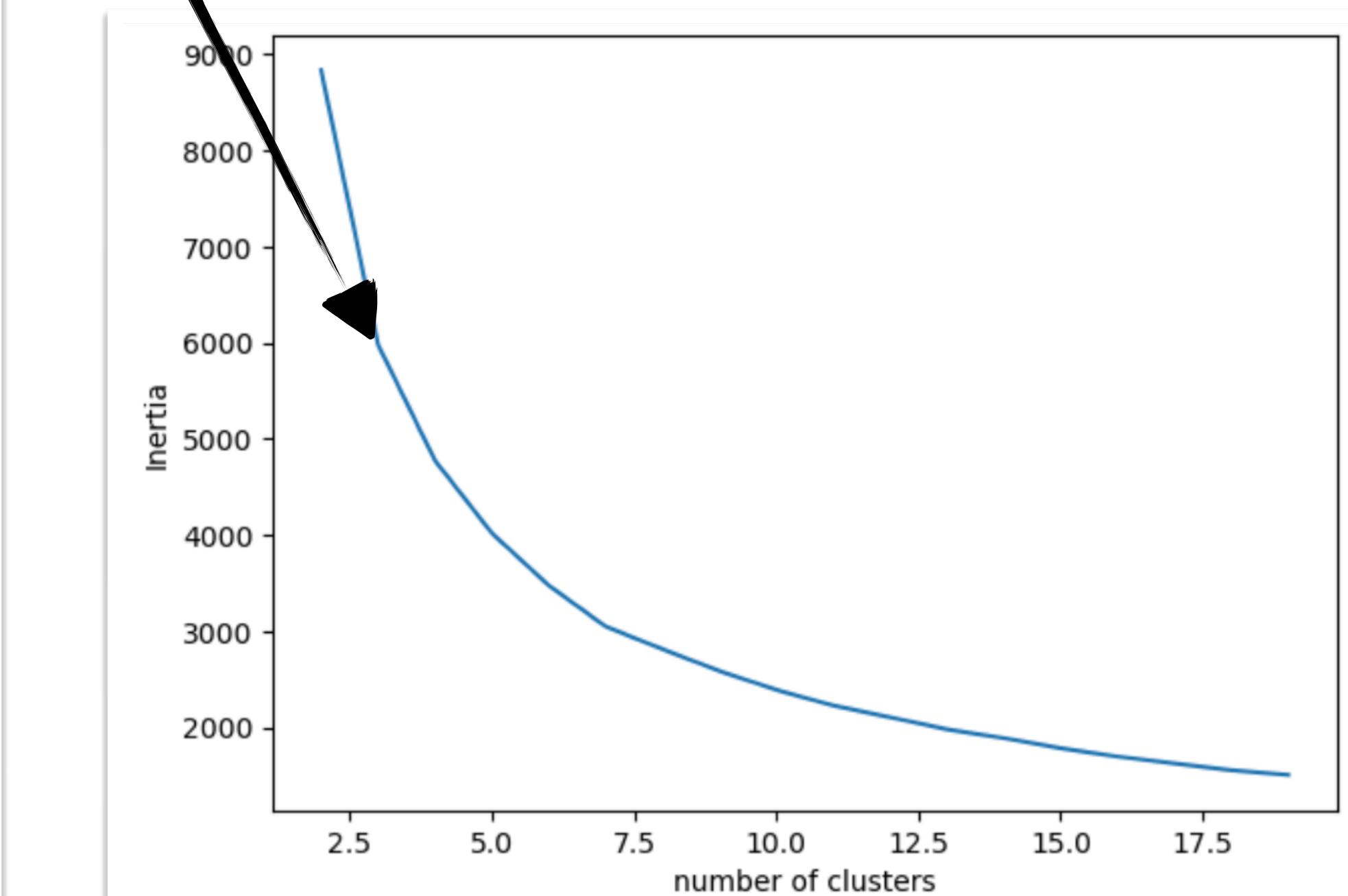
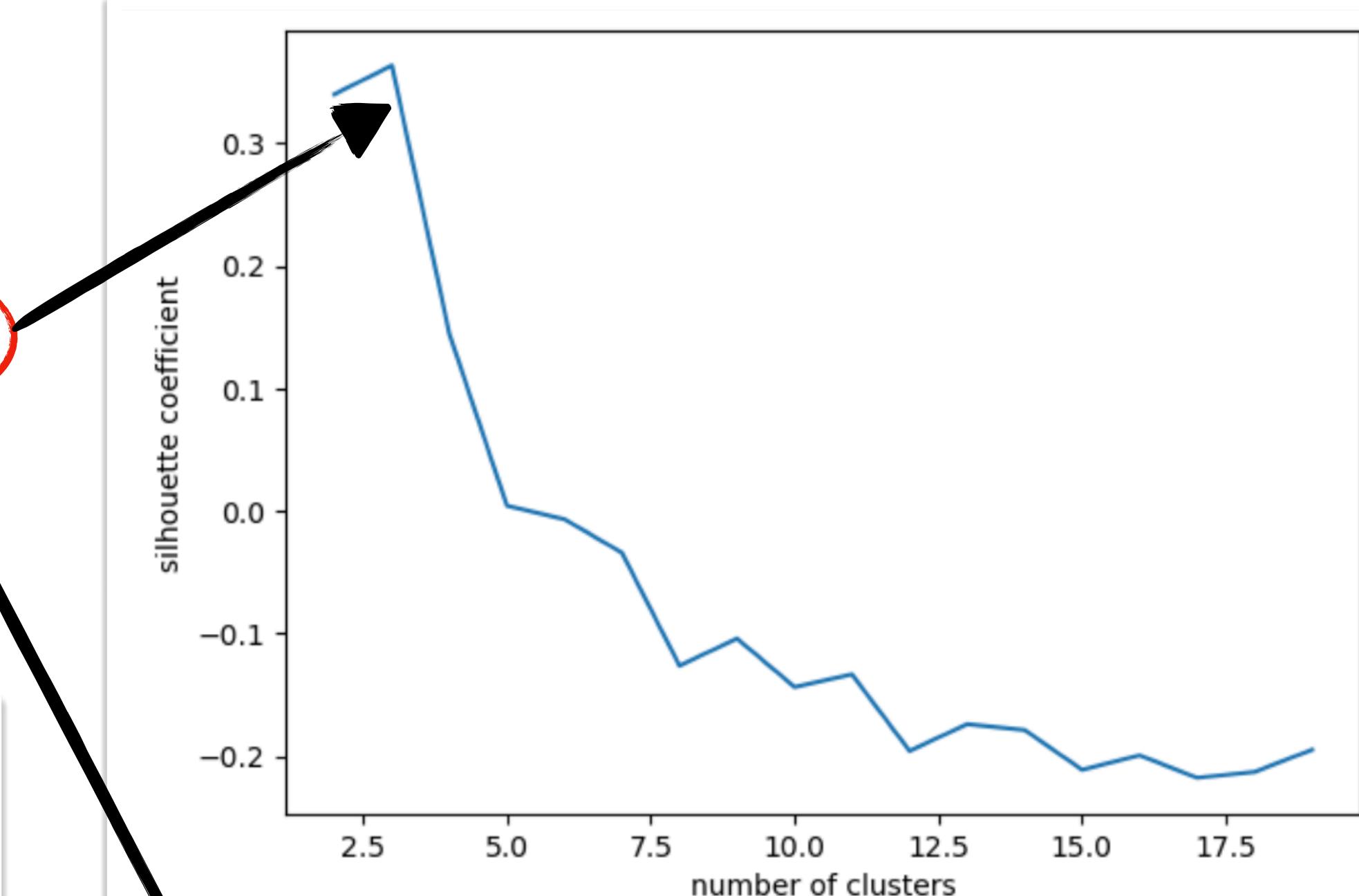
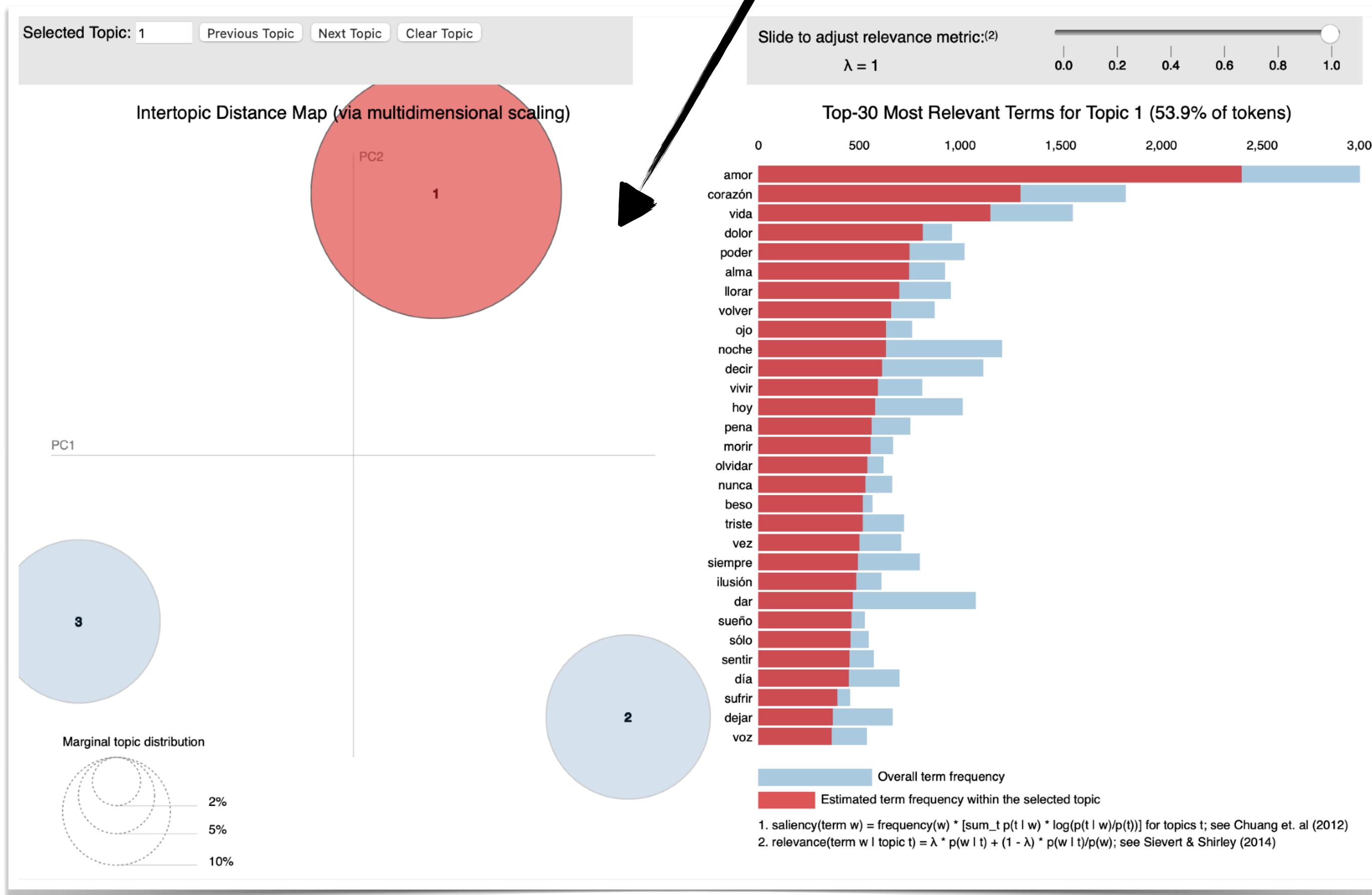


- Cleaning, **preprocessing and feature extraction of the lyrics** dataset, including removal of capital letters and punctuation
- Removal and customisation of **Spanish stopwords** using NLTK
- **Lemmatization** of the lyrics using spaCy
- Model implementation and evaluation using various techniques such as **LSA, NMF, LDA, KMeans, pyLDavis, and PyPI**

sos ojo beso hombre pensar amar buen cariño mujer buscar  
vez pena canción sueño olvidar viejo dar pobre dejar hoy morir  
pasar volver poderasí decir sola alma luz ver vos recuerdo hacer  
tene noche vida querer día corazón si amor

# Results

- 3 topics



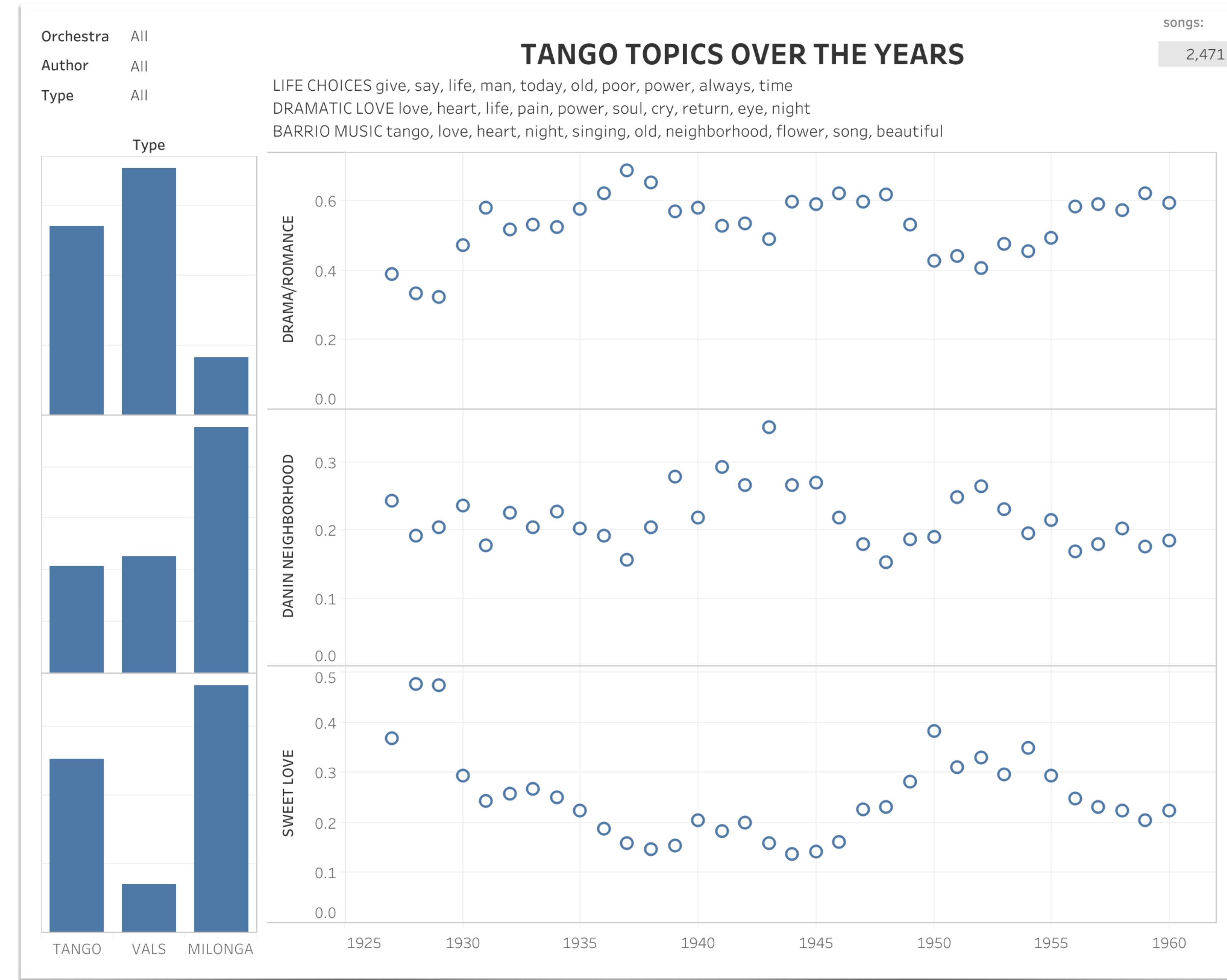
# Results

- **LDA: Latent Dirichlet Allocation**
- **LIFE CHOICES:** give, say, life, man, today, old, poor, power, always, time
- **DRAMATIC LOVE:** love, heart, life, pain, power, soul, cry, return, eye, night
- **BARRIO MUSIC:** tango, love, heart, night, singing, old, neighborhood, flower, song, beautiful

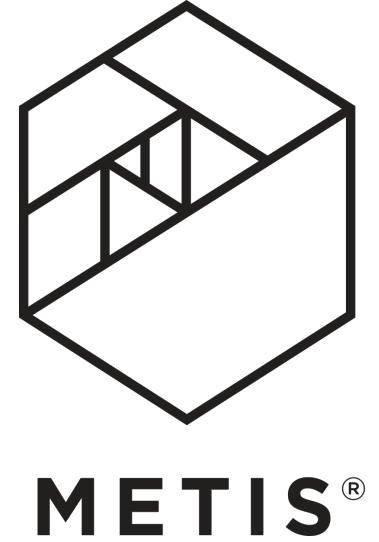


# Results

- Interactive Tableau
- Topics over the years, types, poets and orchestras



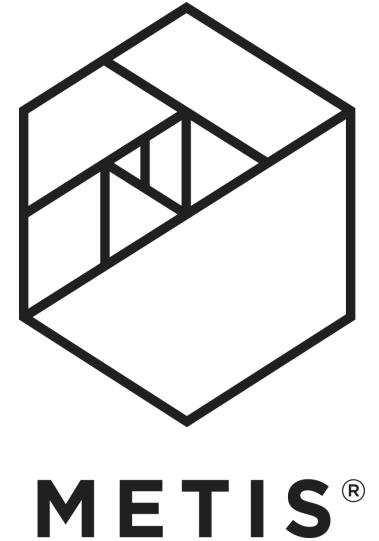
# Conclusions/Recommendations



- **LDA** model effectively identified key topics in Golden Age tango lyrics
- Further analysis of results using **Tableau** for deeper understanding of lyrics by period and orchestra
- Constructed models can serve as a base for **similar explorations**
- Dance and listen to **tango**

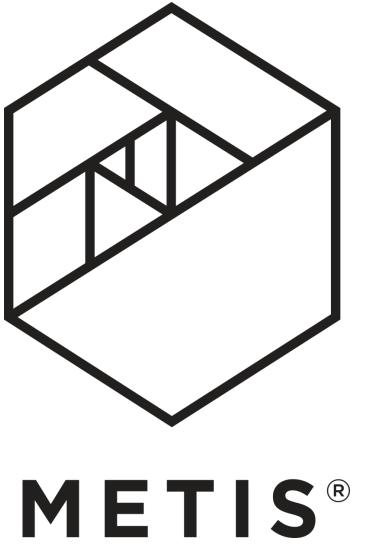


# Future Work



- Utilize cosine similarity for **song recommendation system**
- Explore **other sentiment analysis tools** such as BERT
- Use **more NLP and translation tools** for improved lyrics analysis
- Investigate **text-to-image** technology for live **music visualisation**
- **More data.** Include **music files** for analysis of orchestras and identifying success factors



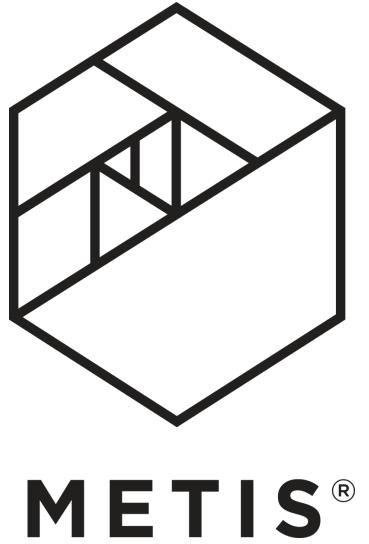


**Thank You!**  
**Questions?**

**METIS Data Science and Machine Learning Bootcamp**

**by Krystian Krystkowiak, 2022**

# Appendix



- I used a dataset from [www.el-recoodo.com](http://www.el-recoodo.com), which is a powerful source of information about Argentinian Tango music for both professionals and hobbyists.
- The dataset was scraped and initially cleaned by gefero, who shared it on GitHub: [https://github.com/gefero/tango\\_scrap](https://github.com/gefero/tango_scrap)