

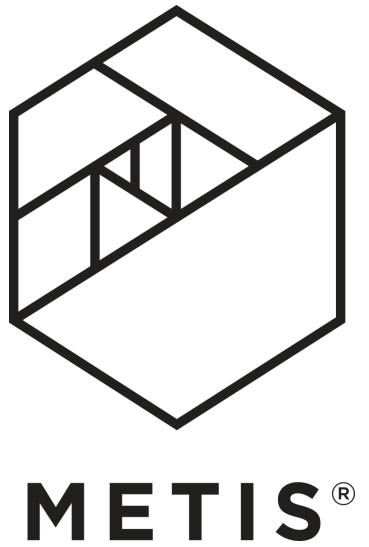
# What are tangos about?

natural language processing for tango lyrics analysis

project for Metis EDA Bootcamp

by Krystian Krystkowiak, 2022

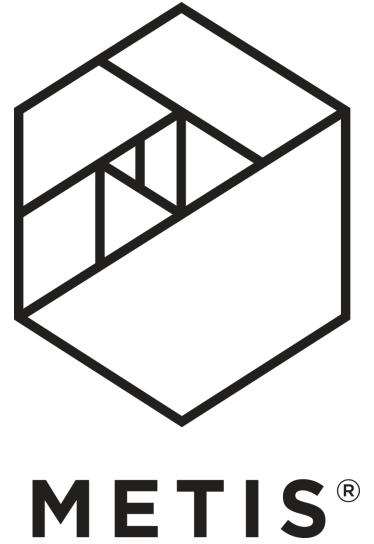
# Introduction



- **Golden Age of Tango Music** - 1935 to 1955.
- **2022 - tango dance is back!**
- But...the **music is old**, from the Golden Age.
- GOAL: **Analysis of Golden Age topics to help modern orchestras** to keep spirit of the period.
- Collectors, teachers and dancers.



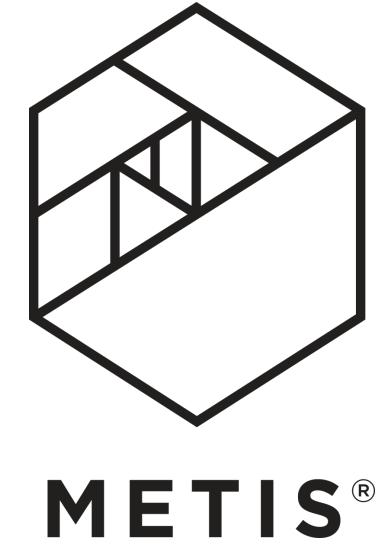
# Methodology



- Dataset from [www.el-recoodo.com](http://www.el-recoodo.com) (scraped and initially cleaned by gefero, who shared it at GitHub)
- Songs from years **1920-1960** to include all traditional tango. **Tango, valses** and **milongas**. The data consist of over **2400 songs in Spanish** with average **145 words**.
- Features: **lyrics** (NLP), name of the song, author, **orchestra**, singer and **year**.



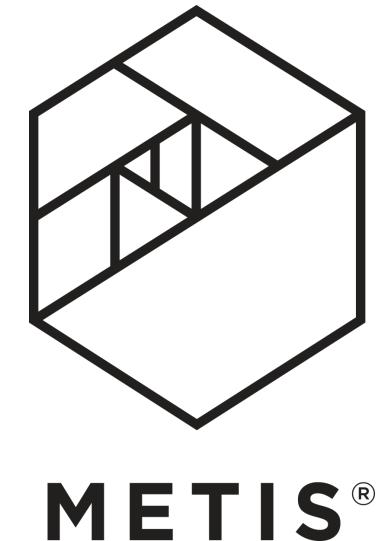
# Methodology



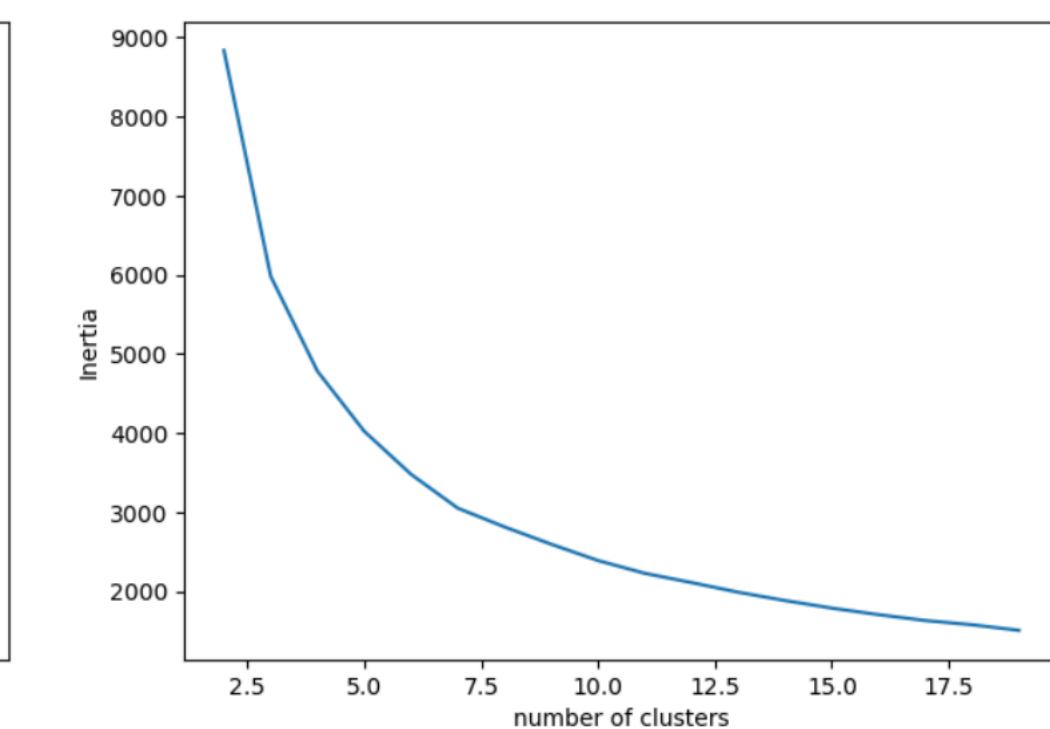
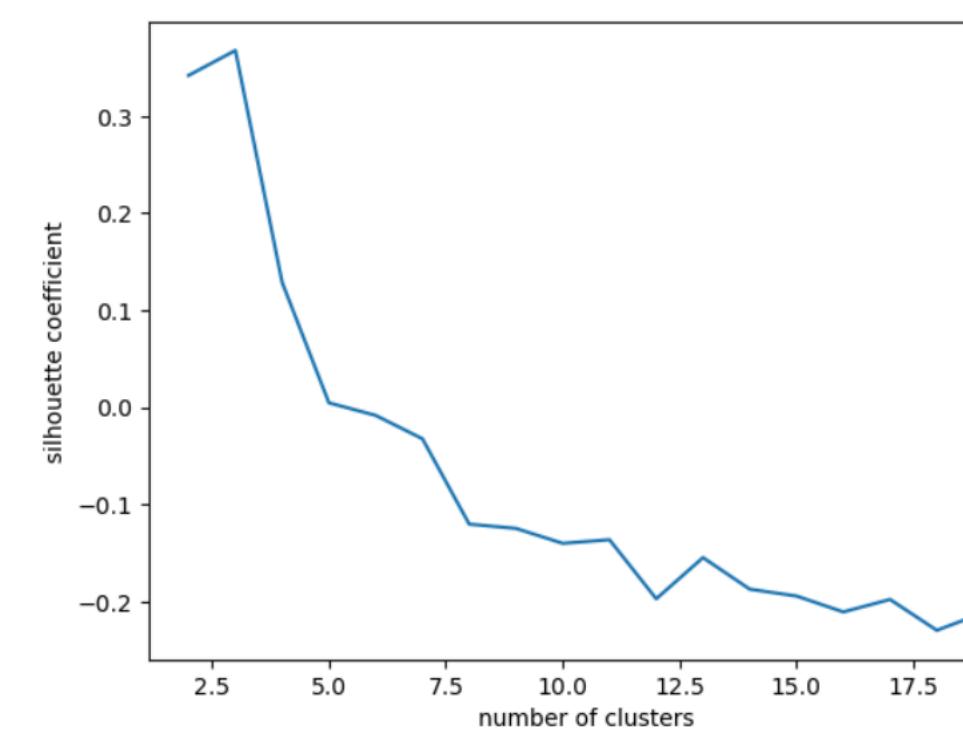
- **Cleaning**, feature extraction, removing capital letters and punctuation.
- **Spanish and custom stopwords** (nltk).
- **Lemmatization** (spacy).
- **CountVectorizer** with couple different models tested: **LSA, NMF, LDA, KMeans, PyPi**.

sosojobarriobeso hombre pensar amar buencariñomujer buscar  
vez penacanción sueño olvidar viejo dar pobre dejar hoy morir  
pasar volver poderasí decir sol alma luz ver vos recuerdo hacer  
tener noche vida querer día corazón si amor  
vivir si prostrarse vez cantar

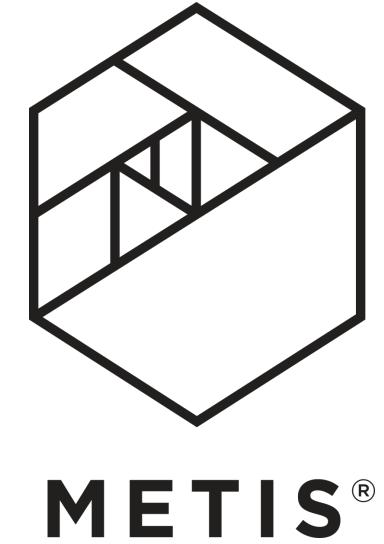
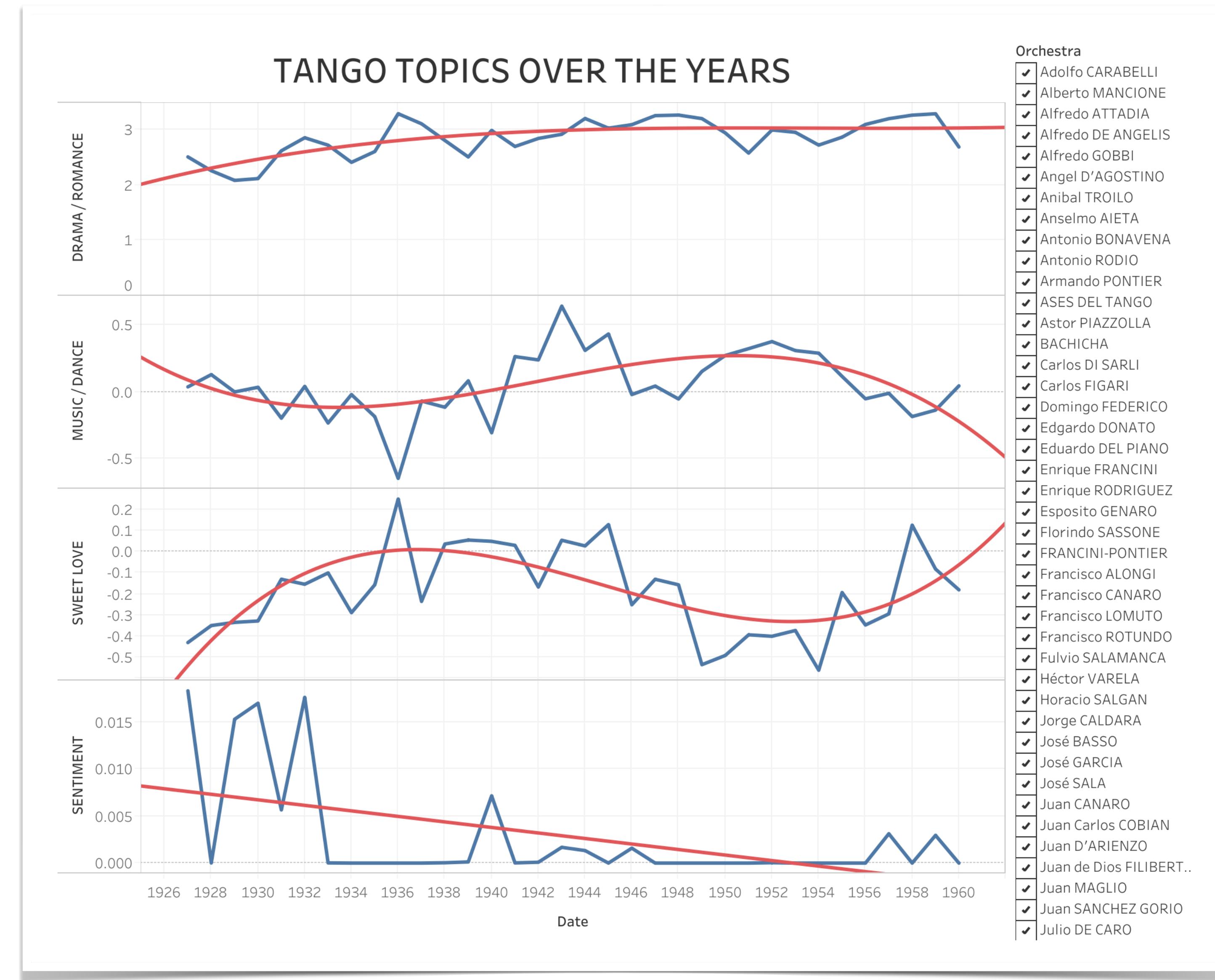
# Results



- **Latent Semantic Analysis** - best results, 3 topics.
- **DRAMATIC ROMANCE** - love, heart, life, night, power, today, say, cry, pain
- **TANGO MUSIC&DANCE** - tango, dance, bandoneon, old man, night, neighborhood, suburb, friend, emotion
- **SWEET LOVE** - love, tango, heart, song, voice, dream, bandoneon, dance, sweet
- **KMeans clustering:**

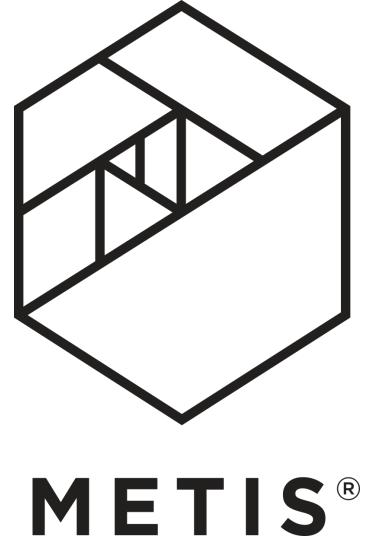


# Results



- Interactive Tableau. Topics and sentiment over the years and orchestras.

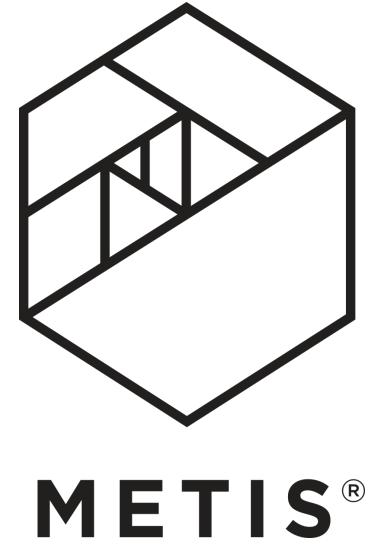
# Conclusions/Recommendations



- **LSA** for shorter texts. LDA did not give satisfying results.
- Deeper analysis of created tableau for **better understanding** of lyrics depending on period and orchestra.
- Constructed models are **ready base for similar explorations**.
- Dance and listen to **tango**.

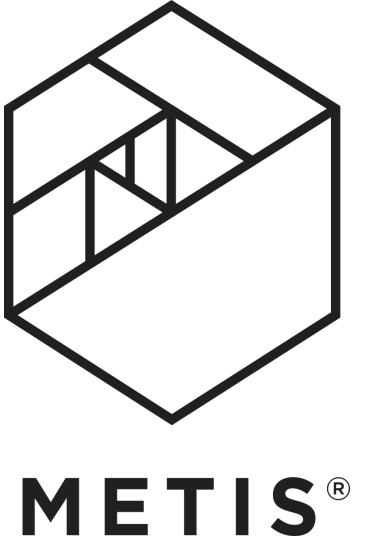


# Future Work



- **Diferent sentiment module.**
- **Exploration of translators** to use more advanced tools.
- **Deeper NLP**, using more tools (translating tangos to emoticons?).
- **Basic recommendation system?**
- **More data.** Music files to find orchestras and songs “success factor”.





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

project for Metis EDA Bootcamp

by Krystian Krystkowiak, 2022