

# Ancient Texts NLP & Classification Project

## Insights from Exploratory Data Analysis (EDA)

All code at [https://github.com/izsolnay/Ancient\\_NLP](https://github.com/izsolnay/Ancient_NLP)

### ISSUE / PROBLEM

Can modern tools provide insights into ancient text corpora?

**Objective:** develop a reliable machine learning model which can classify texts into genres

#### EDA Steps:

- Organize, clean, explore, and analyze dataframe
- Create visuals

### RESPONSE

- Focused on the key variables 'A\_category,' 'B\_category,' and 'God'
- 'B\_category' is best target variable because it contains the genre labels
- 'God' and 'Person' variables make for interesting research target variables. Can texts be classified by god? By person?
- Began initial cleaning such as syncing column names, removing white spaces and removing empty observations

### IMPACT

Results of EDA indicate a severe imbalance by genre ('B\_category'). Some genres have only 1 or 2 texts attributed to them. This is the same for 'God'.

Next steps:

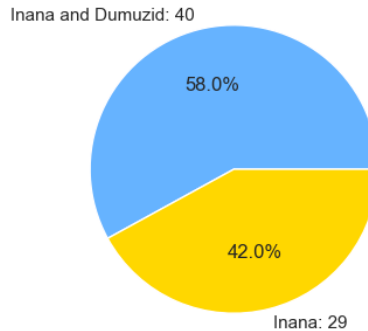
- Address the unbalanced data set
- Investigate and clean text
- Perform sentiment analyses to better understand texts

### UNDERSTANDING GENRE DISTRIBUTION BY TOP GODS

Texts in the data set are categorized into 27 different genres

- Texts feature 1 of 51 gods
- The gods most featured are: Inana, Nana-Suen, and Ninurta

Distribution of Inana with and without Dumuzi in texts



#### Inana

- 12 Royal praise poetry and hymns to deities on behalf of rulers
- 11 Hymns addressed to deities
- 5 Narratives featuring deities
- 1 Hymns addressed to or concerning temples

#### Inana (and Dumuzi)

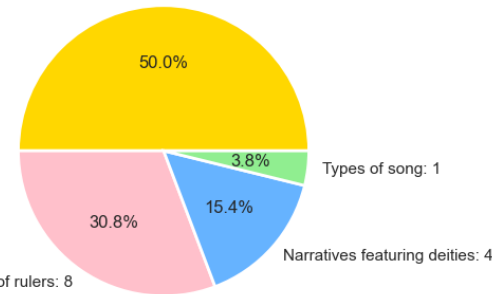
- 35 Hymns addressed to deities
- 5 Narratives featuring deities

Distribution of genre for Ninurta in texts

#### Ninurta

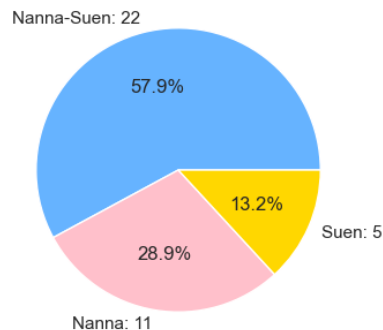
Ninurta is the only god to feature in the genre 'Type of song'

Hymns addressed to deities: 13



Royal praise poetry and hymns to deities on behalf of rulers: 8

Name distribution for Nana-Suen in texts



#### Nana-Suen: name distribution

##### Nanna-Suen

- 21 Hymns addressed to deities
- 1 Narratives featuring deities

##### Nanna

- 10 Royal praise poetry and hymns to deities on behalf of rulers
- 1 Other letters and letter-prayer

##### Suen

- All 5 are Royal praise poetry and hymns to deities on behalf of rulers

### KEY INSIGHTS

The EDA revealed a heavily skewed data set with certain genres far more represented

- 318 out of 555 texts had a hymn or praise label

### QUESTIONS TO ANSWER

- Do different genres exhibit different polarities and subjectivities?
- Do different god-groupings exhibit different polarities and subjectivities?
- Do different person-groupings exhibit different polarities and subjectivities?

