Ancient Texts NLP & Classification Project

Results of Sentiment Analysis
All code at https://github.com/izsolnay/ancient_NLP

ISSUE / PROBLEM

Can modern tools provide insights into ancient text corpora?

Objective
Get polarity and subjectivity scores for texts. Focus on genres and gods.

Sentiment steps
➢ Prepare text
➢ Perform analysis
➢ Create visuals

RESPONSE

➢ Performed basic cleaning & regex formatting of text

➢ Americanized certain words and removed many extraneous terms

➢ Used 2 methods for TextBlob analysis: aggregation of sentences and whole texts

➢ Ran Vader on whole text

➢ Created boxplots for variable genre (`'B_category'`) and Gods.

➢ Created high and low score tables

IMPACT

Results of the analyses indicate that some texts are highly fragmentary and that there is a severe imbalance by genre

Next steps:

➢ Address the unbalanced data set

➢ Investigate and clean text

➢ Get words counts and weights to better understand sentiment scores

UNDERSTANDING SENTIMENT
Positivity, negativity, and subjectivity

Method 1
Sentences aggregation

• No text scored above a 50% positivity rating
• The gods Nana, the Ekur, Nisaba, Inana and Dumuzi feature in highest rated texts

Method 2
Whole text aggregation

• Texts had much higher positivity
• *A song of Shulgi* had the highest positivity rating at 84%
• *A drinking song* was the 5th highest rated in positivity

Ratio of positive texts to all texts by featured gods

God	percentage_highscore	percentage_lowscore	Count
Inana and Dumuzid	50.0%	50.0%	32
Inana	36.0%	64.0%	25
Gods	8.3%	91.7%	24
Nanna-Suen	65.0%	35.0%	20
Ninurta	42.1%	57.9%	19

KEY INSIGHTS

➢ Whole text analysis is the superior method
➢ Fragmentary nature of many sentences skews the aggregation method results negative

QUESTION TO ANSWER

Which words and phrase are common to which genres, gods, and persons?

THIS WILL

Prepare for machine learning and reveal key insights for how sentiment scores were informed