

Features:

There are 8265 features, all of which are words from the religious texts.

Prompt:

1. Look and explore through the data. Organize your findings and observations in a way that will help you visualize certain patterns.
2. Extract trends and patterns from the data using:
 - a. Data visualization: Create an infographic with no less than three charts (that can help you better understand the data).
 - b. Hypothesis/experimental testing: Explore the data and come up with a hypothesis.
3. Use data visualization tools (Python libraries, Tableau, etc.) to present the trends you found. Create an infographic with no less than three charts.
4. Using your findings, determine:
 - a. the top 20 words that appear in each book
 - b. the top 20 words that appear through all books
5. Three of the books (Proverbs, Ecclesiastes, Wisdom) are part of the Old Testament, but are spread out in time of founding. Define a model to determine how the wording has changed from the Book Of Proverbs to the Book of Wisdom. This is an open-ended question, so you may choose to answer it in any manner appropriate, as long as you use a machine learning method.
6. Both Buddhism and Taoism have influenced each other. Propose a hypothesis to find out similarities that can be drawn between both texts in terms of word usage? Can we extend the word usages in the texts to their practices? Justify.
7. Prepare a report containing your results from the analysis. It should contain the following: Intro, data cleaning/pre-processing, visualizations (at least 3), analysis, proposal, conclusion.