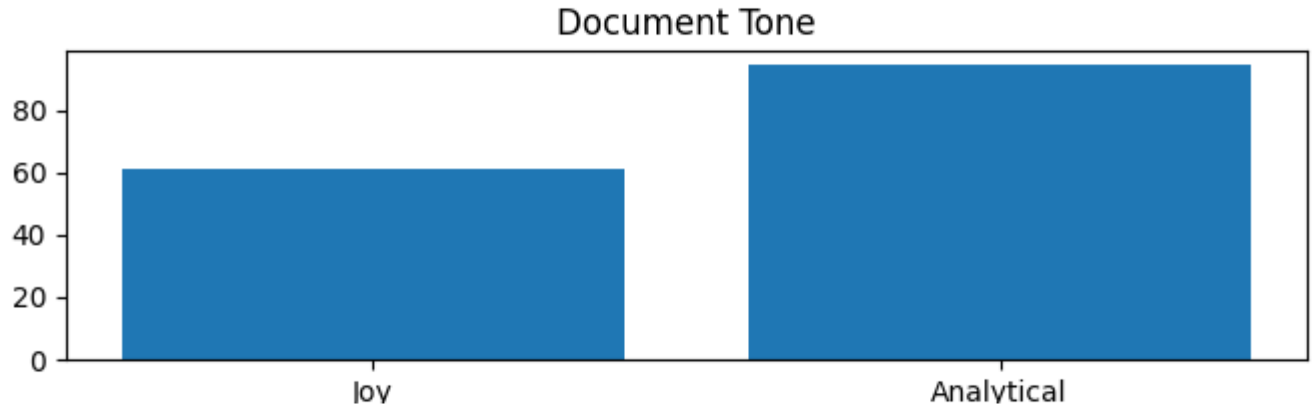
Original Text

Social Scientists especially the historians are very much interested in analyzing historical documents from various sources in order to generate lesson-oriented concrete and objective analysis of the past. Such documents, however, contain mostly the textual data and can be in various formats (e.g. pdf, csv, scanned images etc). The textual nature of document and their divergent formats make it quite tedious for social scientists to highlight or de-emphasize certain angles in order to generate the analysis in their own line of understanding. This project aims to develop a toolkit that allows them to feed documents of various formats and provides them with a useful set of data science tool (Topic Modeling, Data Mining, Text summarization, Sentiment Analysis) to quickly extract useful information that in turn can be used to form their analysis or to uncover newer insights. In the end the usefulness of the developed toolkit will be demonstrated by applying it to researched textual data of British rule in South Asia from British and south asian perspectives.


Document Tone
Joy
Analytical