Executive Summary Qarik Corporate Project Group Bear

Qarik provides a corpus of World Bank Loans. The goal of this project is to offer insights into economic and development trends from this unstructured data. Qarik would like to see the ways in which we can discover, extract and display these insights from this large corpus of text.

The stakeholders for this project consist of the World Bank itself, economic analysts, and governmental bodies applying for loans. The key performance indicators we are keeping in mind are to successfully extract over 90% of the data for each feature, bring some structure to our data via clustering, provide some analysis of economic trends in relation to the loan agreements, and provide visualizations to provide explainability.

Story: TBD

Group Bear hopes to achieve this goal by first extracting the text in each Loan Agreement and then cleaning the data to gather features of interest. In particular, we found that the country, date, loan amount (and currency), project name, and project description to be features of interest. After extracting the relevant features from the corpus provided, we also obtain additional economic data on the countries in the loan amount in order to aid in our analysis of different economic and development trends. In particular, we provide an analysis of GDP for particular countries to see any trends for these countries that have obtained loans from the World Bank. A visualization in Tableau is provided on a world map to show global relationships between features for the whole corpus for a non-technical audience.

In order to provide further structure to the data, we provide two models (LDA and k-means) for clustering the corpus by which sectors of the economy are impacted from the project descriptions and project names for each Loan Agreement. The value of these models is to compare with the World Bank’s sectors for the loan. Some projects are an amalgam of sectors and the LDA model shows these relationships in its clustering. The k-means clustering uses a GloVe model to represent documents as vectors and determines a clustering based on similarity to keywords provided by the World Bank’s description of each sector.

Conclusion: TBD