DATS 6103 Final Project Proposal:

Group 2

Lobbying activities are crucial in influencing policy and governmental decisions nowadays. It is essential to know the trends, patterns, and dynamics of lobbying expenditures and activities to understand how interest groups, industries, and policymakers interact. Our final project aims to provide a multifaceted analysis of lobbying behaviors and expenditures over time including industry involvement and relationships between policy areas.

Our dataset focuses on analyzing Senate lobbying disclosures obtained through the Lobbying Disclosures Act API (https://lda.senate.gov/api/v1/filings/). It includes quarterly filings with lobbying registrant/client information. We will use paginated queries to make sure that all the data is collected.

With this dataset, we will have three main potential research questions, with the final research questions depending on the sufficiency of our dataset.

1. How have lobbying expenditures changed over the last year, and are there any notable trends over this period? This will likely be conducted through either linear regression or time series analysis.

2.1 How do lobbying efforts (general expenditures, individual lobbyist allocations) differ between policy areas in 2023? The analysis method we will use is K-means clustering.

2.2 How does the likelihood of policy area focus differ by the size (in total contributions, total lobbyists hired, etc.) of the lobbying client and/or registrant? Testing will be done using the logistic regression approach; we may also attempt to examine probability through K-means clustering with binary variables. These two research questions aim to explore which industries are most actively involved in lobbying activities and differences across different years and filing periods.

3. For a given policy area, what is the likelihood that lobbying efforts would be conducted jointly with other policy areas? Put differently, which policy areas have the closest associations? This will be conducted through either a correlation test or a Chi-square Test; we may also attempt to utilize the apriori algorithm approach to try and identify such associations

git repo link: https://github.com/Xue-Ming-Wang/DATS-6103-final-project.git