

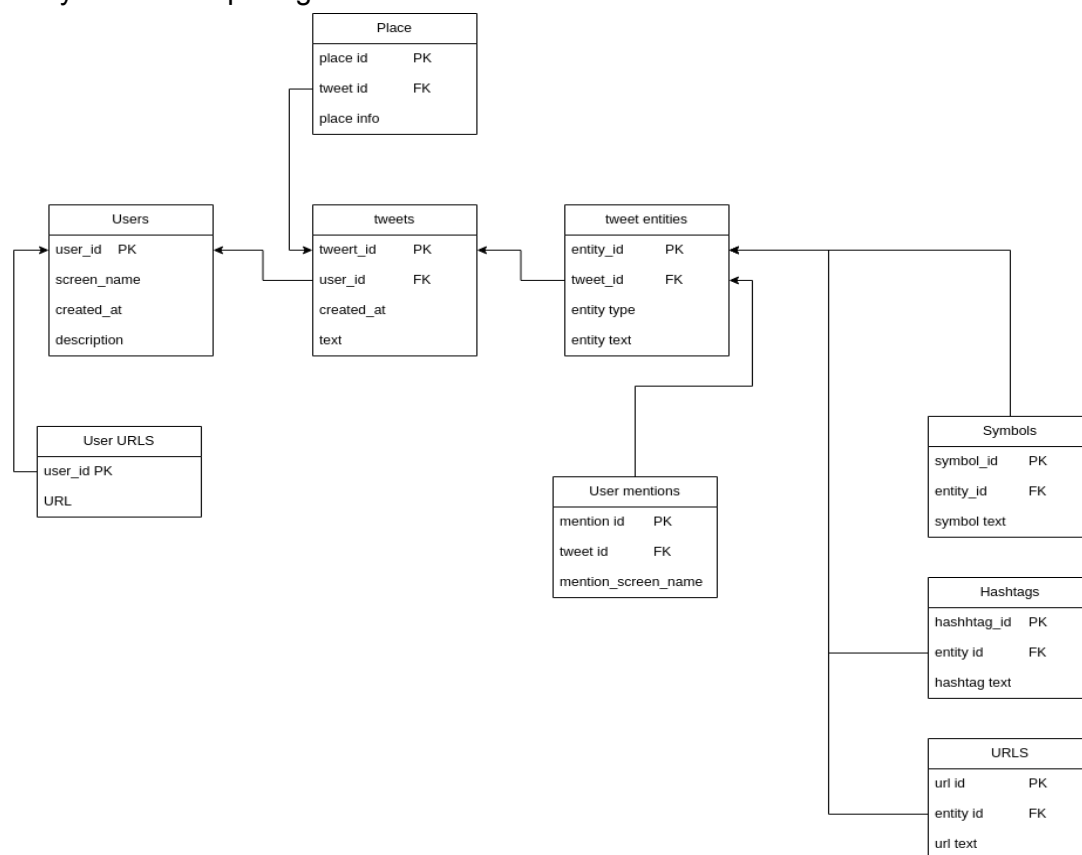
Project Proposal for Lobbyists4America: Analyzing Congressional Tweets (2008-2017) for Lobbying Insights

Description: The project aims to analyze congressional tweets from 2008 to 2017 to gain insights into key topics, influential members, and relationships within Congress. The findings will be of interest to Lobbyists4America, a company seeking to strengthen their lobbying efforts by understanding legislative trends and member engagement on Twitter. The audience for this project includes lobbyists, policymakers, and individuals interested in understanding the dynamics of congressional discussions on social media.

Questions:

1. What are the most discussed topics in congressional tweets during the specified period?
2. Which members of Congress are the most active on Twitter and engage with legislative topics?
3. Are there any significant connections or relationships between members based on their interactions on Twitter?

Entity Relationship Diagram



Hypotheses:

1. Congressional tweets will cover a wide range of topics, including healthcare, economy, defense, and immigration.
2. Influential members of Congress, committee chairs, and leaders are expected to be more active on Twitter and have higher engagement with tweets related to legislation.
3. Social network analysis might reveal clusters of members who frequently interact or retweet each other, indicating potential alliances or affiliations.

Approach: The initial approach involves data collection, importation, and cleaning. We will then perform initial exploration to understand data distributions and patterns. Key features to be analyzed include tweet text, timestamp, user mentions, and retweet counts. We will apply natural language processing techniques to extract topics from the tweet text and identify frequent co-occurrences among members.

For the analysis, we will conduct social network analysis to uncover relationships and connections between members based on interactions. Additionally, sentiment analysis will be used to gauge the tone of congressional tweets over time. The primary evaluation metric will be the identification of significant topics, highly engaged members, and clusters of interconnected members.

The project will be iterative, and we may modify the approach as we uncover new insights during the analysis. Collaboration with peers on the discussion boards will help brainstorm ideas and gain diverse perspectives on the data and its potential interpretations.