Liam Gersten
Pittsburgh, Pennsylvania
lgersten@andrew.cmu.edu
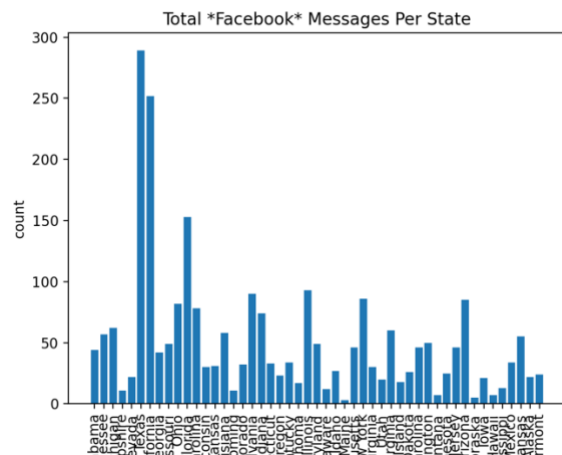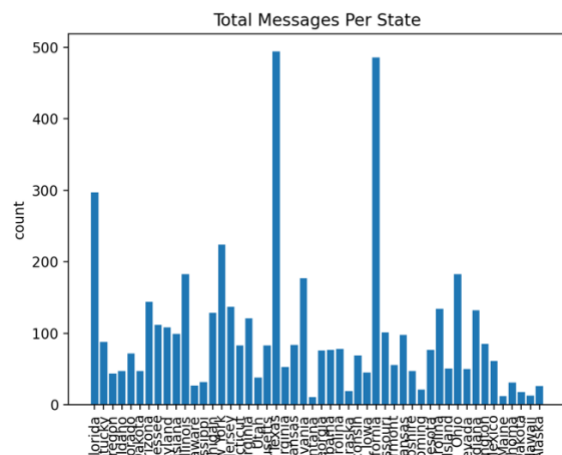
Carnegie
Mellon
University

Summary of dataframe:

- Parsed and cleaned dataframe, df includes 4,980 entries, with 27 columns.
- Of the 27 columns, 9 include categorical entries, 5 include quantitative entries, and 2 include lists.
- Additional entries are contained in separate dictionary variables in printOuts() and displayFigures().
- The printout for df.info() is as follows:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4980 entries, 0 to 4979
dtypes: bool(1), float64(6), int64(2), object(18)
Data columns (total 27 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   unit_id               4980 non-null   int64
 1   golden                4980 non-null   bool
 2   unit_state            4980 non-null   object
 3   trusted_judgments     4980 non-null   int64
 4   last_judgment_at      4980 non-null   object
 5   audience              4980 non-null   object
 6   audience_confidence   4980 non-null   float64
 7   bias                  4980 non-null   object
 8   bias_confidence       4980 non-null   float64
 9   message               4980 non-null   object
 10  message_confidence    4980 non-null   float64
 11  bioid                 4980 non-null   object
 12  embed                 4980 non-null   object
 13  id                    4980 non-null   object
 14  label                 4980 non-null   object
 15  message_gold             0 non-null   float64
 16  source                4980 non-null   object
 17  text                  4980 non-null   object
 18  name                  4980 non-null   object
 19  position              4980 non-null   object
 20  state                 4980 non-null   object
 21  region                4980 non-null   object
 22  hashtags              4980 non-null   object
 23  date                  4980 non-null   object
 24  numeric_date          4980 non-null   float64
 25  sentiment             4980 non-null   object
 26  exact_sentiment       4980 non-null   float64
```
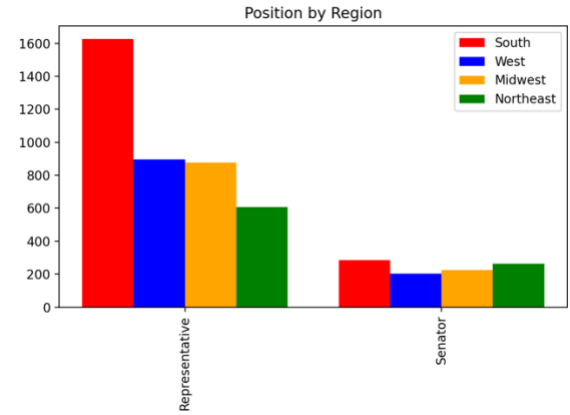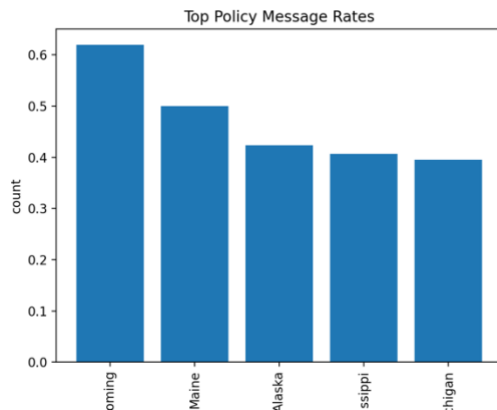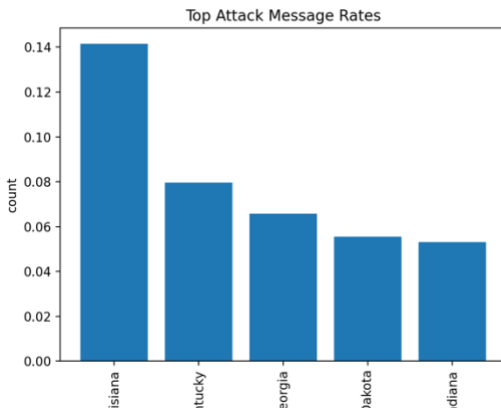
Summary of printouts:

- Total Messages Per State: dictionary mapping states and message frequency, sorted by frequency.
- State Counts for Negative Sentiment: sorted dictionary mapping states and negative sentiment frequencies.
- Message Types for Region: dictionary mapping regions to sorted dictionaries that map message types to relative frequency.
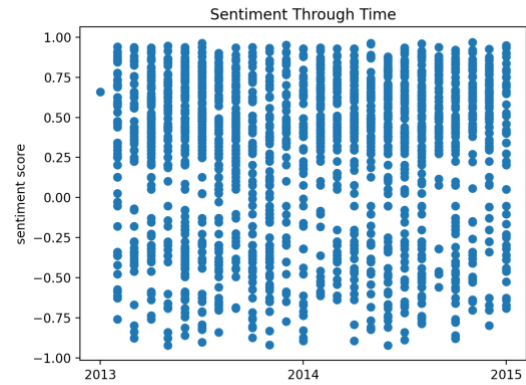
Bar charts:

Filtered bar charts:

Top Attack Message Rates

Top Policy Message Rates

Side-by-side bar charts:

Messages by Region

Position by Region

Scatterplots:

Sentiment Through Time

50 common hashtags by frequency and sentiment