Analysis Data File: Scored_Data.csv

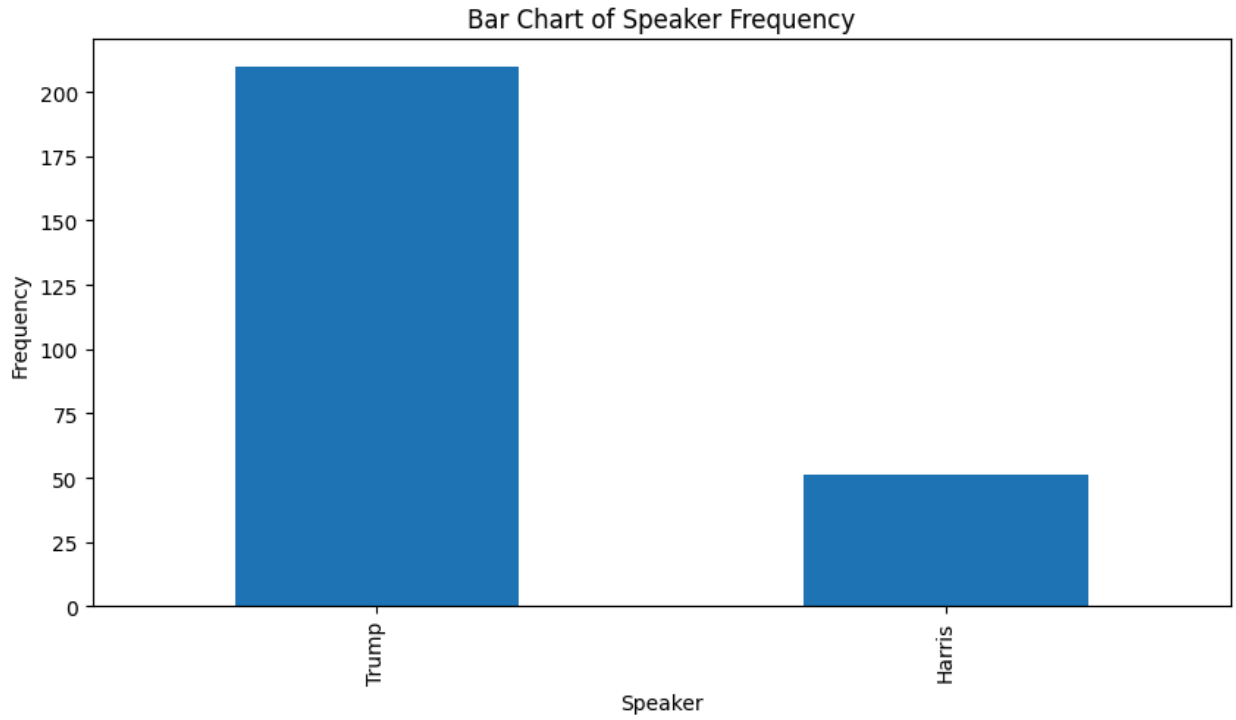
Unit of observation: Comment said by a candidate

Variables

Name	Definition and Source	Number of Observations (Missing)
Speaker	Which political candidate made the comment. Automatically assigned to comments drawn from Harris_Philadelphia_Transcript.txt and Trump_Wisconsin_Transcript.txt, assigned to comments from Debate_Transcript.txt based upon who made the comment.	261 (0)
Topic	Overarching policy issue being discussed. Column populated by ChatGPT, which iterated through data in the 'Speech' column and assigned a Topic from this list: criminal justice, economy, health care, abortion opposing candidate, education, immigration, war in Middle East, war in Ukraine, or none.	261 (113)
Score	Sentiment score generated by VADER sentiment analysis, ranging from -1 to 1, based upon data in the 'Speech' column.	261 (0)
AtDebate	An indicator of whether the comment was made at the debate or not at the debate (at a rally). This was automatically assigned to each entry based upon which raw data source it came from.	261 (0)
Speech	The comment made by the candidate about the topic. This data comes from each chunk of text in each of the three raw transcripts.	261 (0)

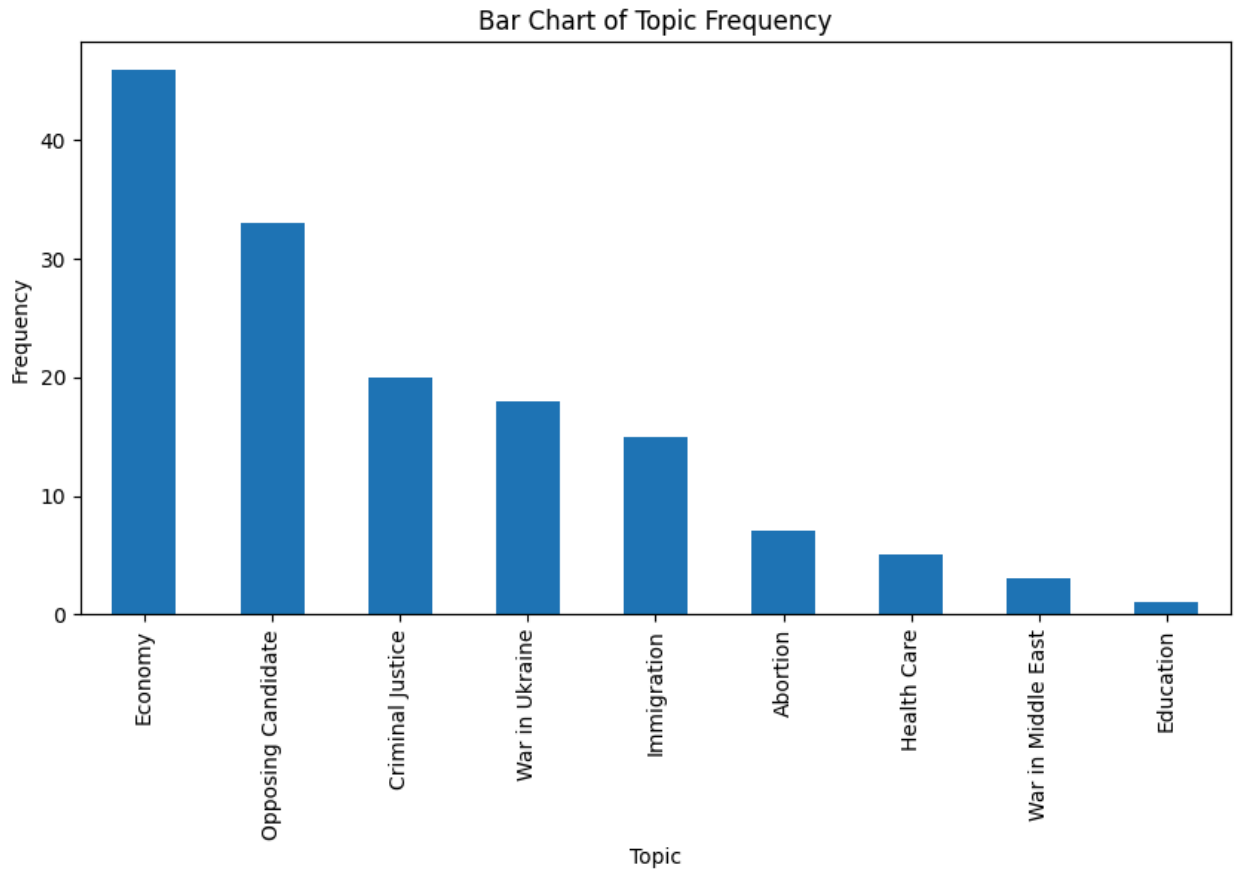
Variable Data
“Speaker”

Value	Count
Trump	210
Harris	51



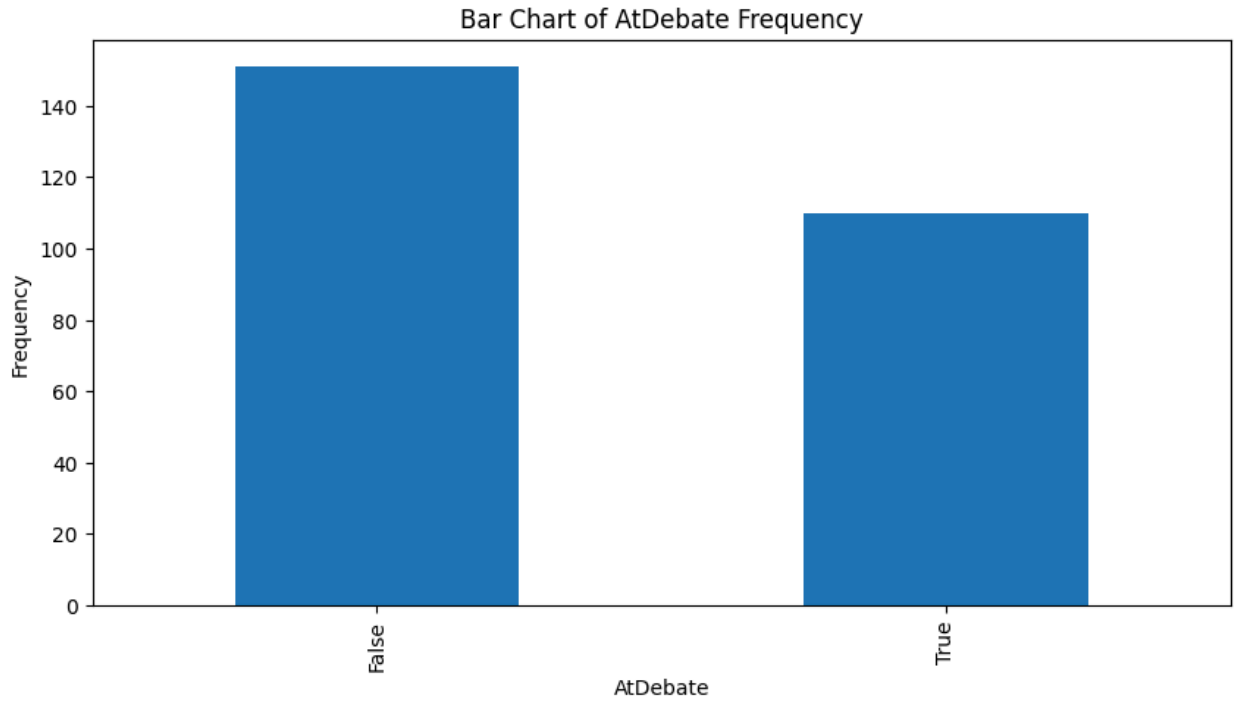
“Topic”

Value	Count
Economy	46
Opposing Candidate	33
Criminal Justice	20
War in Ukraine	18
Immigration	15
Abortion	7
Health Care	5
War in Middle East	3
Education	1



“AtDebate”

Value	Count
False	151
True	110



“Speech”

For this data, we will not provide any additional information. The tables and diagrams associated with categorical and quantitative data would not fit this variable, as it consists of 261 unique text entries.

“Score”

Statistic	Value
Count	261
Mean	0.0718
Standard Deviation	0.7538
Minimum	-0.9964
25th percentile	-0.7633
50th percentile	0.0772
75th percentile	0.8850
Maximum	0.9962