# Research Question:

Do publications negatively portray certain candidates more than other candidates?

## Introduction

A news organization’s role in society is to gather, write and distribute key information to the masses. Depending on the new organization’s specialty, its journalists aim is to keep the public informed on foreign and domestic issues and events. Given its expansive audiences, news media organizations have long been utilized to influence public perception on important issues and help propagate many political initiatives.

Freedom of the press, the first amendment in the United States Constitution, is a key pillar in American democracy. The amendment and has played an important role in politics since its creation as its intended purpose is to help the American people make informed decisions without government intervention. Thus, the role of American news organizations is to help cultivate the opinions of the American society. Given this power, a news organization plays a significant role in each election cycle based on how it covers each candidate and political party. This project aims to assess whether news organizations have abused this influence by negatively portraying certain candidates in the United States 2020 presidential primary.

During the presidential primary, a news media organization tailors the public’s perception of the candidates through its delivered content. The content could possess a certain level of bias toward candidates due to factors such as frequency of coverage, data journalism, connotation, verbiage, and image choice.

The frequency of coverage alone can greatly affect society’s perceptions on a candidate or political party. Name recognition is cited as having a very large influence on each election and respective primary. When news coverage is mainly centered on a singular or selection of candidates, other candidates and their proposed policy changes can be virtually invisible to the American people. Being for-profit institutions, disproportionate candidate coverage may have more to do with revenue than intentional institutional bias of its journalists. Clicks, up votes, comments, re-tweets, citing and republication tends to favor the more polarizing candidates.

Catering to the partisan audience is another way a news organization can influence its readership during an election period and may affect how a news outlet portrays certain candidates and political parties. Based on a study by the Paw Institute, the perceived political slant of a news organization aligns with the political affiliation of its consumers. This facilitation of ideological bias by its consumers may influence how a story is politicalized and sensationalized by the new organization’s employed journalists. Bias can become a self-perpetuating system, sustained by both the consumers and the new organizations. During an election or primary, this partisan coverage can force a news organization to place emphasis on the candidates that are winning, losing, or posing the largest threat to the organization’s desired candidate.

Word choice is another key tool a news organization utilizes to subtly convey its feelings toward the candidates. Words are powerful and are intentionally selected to help evoke a certain feeling resulting from their connotations. For example, a news article may use words with violent connotations to help describe a certain candidate’s proposed policy change. The violent words subtly imply that the constituents would be victimized by the policy change. In conjunction with aforementioned point on the partisan consumer, the audience may not realize or care about the content’s intentionally biased wording.

Similarly, the images that a journalist or editor uses are intentionally selected to fit the connotation of the accompanying article. Visuals can evoke feelings ranging from fear and despair to happiness and acceptance. As people tend to be more visually inclined, the reader can be influenced to feel a certain way about the highlighted issue. The inclusion of candidate visuals for articles is a tactic many news organizations use in elections and primaries to influence its consumers.

Data journalism has been wielded as a weapon by news organizatios, especially during the election periods. During the presidential primary, candidates are frequently asked questions about their proposed policy initiatives and stances on certain issues. As technology has improved, the concept of live fact-checking has come into vogue. The news organizations have started employing these fact-checking techniques to facilitate the jockeying of certain candidates.

In addition to fact-checking, news organizations have historically incorporated polling data into its delivered content. The polling data acts as a mechanism to influence voter perception and can also impact the amount of coverage the candidate receives by the news organization. Due to the nature or media coverage, the inclusion of polling data becomes a self-perpetuating cycle. Media flocks to the front-runners, who in turn receive more media coverage, and consequentially, climb even higher in the polls.

While the intended goal of a news media organization is to help inform the American people, the objectivity of its reporting should be questioned based on the aforementioned. To assess whether news organizations have negatively portrayed certain candidates in the presidential primary, the team will focus on candidate biases created based on word choice. The project will employ a sentiment analysis model to determine the subjectivity and the overall emotional bias of candidate articles from a selection of news syndicates.

In this paper, the Literature Review section focuses on academic research pertaining to new media biases. The data set and approach section reviews the methodologies employed to cultivate and assess the sentiment of media content. The bias design section focuses on model used to measure bias. Lastly, the results section provides a summary of the insights derived from the team’s analysis.

## Literature Review:

The Literature Review section focuses on a selection of published literature and analyses that define political bias and identify ways of measuring and quantifying these biases. The section is divided into three key components: causes of bias, consequences of bias, and the measurement of bias.

### Causes of Bias

Several papers, such as Sutter [1] and Baron [2], focus on the factors that propagate political biases. Sutter [1] compares the incentives for bias in universities versus those in for-profit organizations. He concludes that the forces that sustain bias in academia differ than those in commercial organizations since the stakeholders and accountability structures are vastly different.

For Bernhardt [3], media bias is a result of the general public’s preference for “partisan” or biased news. Baron [2] concludes that news bias is a self-perpetuating cycle where journalistic biases lower subscription prices, forcing the news organization to survive only by hiring and paying journalists lower wages.

Gentzkow et al [5] aligns the causation of political bias to the production and dissemination of news: supplyside and demand-side, respectively. The supply-side biases can be attributed to the production of news media content, such as practices of management or labor in news media organizations. The demand-side biases originate and perpetuate from consumers’ perceptions and preferences.

For bias to persist, there must be factors that sustain, or even strengthen it. The task of identifying factors that sustain bias across the adaptive and dynamic environment of news media organizations is the focus of papers such as Sutter [1], Baron [2], Bernhardt [3], Baum and Groeling [4].

Sutter [1] suggests that journalism, the main transmission engine for ideas and information in society, is inherently subjective and often reflects the opinions of the journalist and/or news outlet. However, the evidence to prove the existence of bias is circumstantial, making the measurement and quantification of these biases very difficult. Sutter notes that revenue is a controlling force for limiting or reducing bias in privately-owned news media organizations. He argues that the inclusion of too much bias in reporting can adversely affect the readership and advertising revenue for the news organizations.

Like Sutter, Baron [2] submits that profit-maximizing, competitive pressures, consumer preferences, and the general economic and financial landscapes impact the role of bias in news media. Unlike Sutter, Baron theorizes that there are strong correlations between political biases and consumer expectations, operational costs, and the ideological alignments of the employed journalists. Baron’s findings show that several forces combine to ensure the presence and prevalence of unbalanced reporting.

Bernhardt [3] also suggests that profit-maximization may influence a news media outlet to cater to a more partisan audience by suppressing or doctoring some pertinent information in its news reporting.

Baum and Groeling [4] submits that the advent of internet media removed the professional editorial process from news content. The new “partisan” styled editorial system of internet news, paired with the public’s adoption of the internet, produced a more fragmented audience.

### Consequences of Bias

Several authors explain the different ways that political bias causes harm. Bernhardt [3] cautions that the impact of media bias is more powerful and determinative than commonly presumed, since its effect cannot be nullified, even to a rational consumer. Baum and Groeling’s [4] study found that left-skewed and right-skewed political biases exist in internet news media. This shift from bipartisan to partisan coverage has important implications for political discourse in America and may impact the decision making for many Americans.

Like Baum and Groeling, Gentzkow and Shapiro [5] believe that media bias can have large effects on voter behavior, public perception and political outcomes. However, it is very difficult to determine the actual impact of these biases.

Stone [6] presents the counterintuitive thesis that bias can be more prevalent in certain competitive media markets than in some less competitive ones. Stone presents formal proof that consumers can, for example, be less informed in some media duopolies than in some monopolies. Thus, one must guard against simplistic prescriptions on how to counter bias, since its correlations with market conditions can be contrary to intuition.

Sheth’s [7] initial findings are that media coverage is broadly anti-conservative. His model also concludes that the news outlets, regardless of political affiliations, are resoundingly negative and critical of the opposition instead of championing its respective candidates. The Sheth model findings align with the general public’s perception of media bias.

Lin, Bagrow and Lazer [8] prove that internet blogs are, more “social” as they are more connected and consequently more influenced by site connections. Being more connected makes the internet blogs more “exogenous”, or more influenced by external factors, such as election cycles.

### Measuring Bias

Many researchers have grappled with the difficulty of precisely measuring or documenting bias. A number of papers, such as Sheth[7], Lin [8], and O’Connor [9], focus on the problem of quantifying media bias in networks, and present models that can be used to quantify political bias.

Sheth [7] proposes a model for measuring ideological bias in news coverage by mainstream media outlets. Terms in the news articles are classified based on detected political affiliations: pro-conservative, anti-conservative, pro-liberal and anti-liberal. Once classified, the model detects the overall measurement of political bias in each article.

Lin, Bagrow and Lazer [8] propose a method of quantifying bias in both traditional news media and internet blogs. The method primarily compares the number of media references to members of the United States congress. This method helps reduce errors due to subjective criteria, such as sentiment or degrees of bias. The presented method can also account for the variance in bias over time.

O’Connor, Balasubramanyan, Routledge and Smith [9] test how well Twitter sentiment correlates with consumer confidence, presidential and election polling data. The highlighted sentiment analysis model uses a lexicon from OpinionFinder on a tailored subset of Twitter data relating to several political topics between the years 2008 and 2009. The topics include key words such as “jobs”, “Obama”, “McCain”, etc. The model calculates a sentiment score based on the ratio of positive to negative words. Smoothing averages techniques remove sentiment score noise and can help to determine correlation. O’Connor, Balasubramanyan, Routledge and Smith note that techniques like stemming caused dangerous misclassification over the model’s lexicon, but did not substantially affect the aggregated scoring composites. The model proved effective with predicting economic indicators, but exhibited difficulty predicting the presidential approval polls.

The model described by Groseclose and Milyo [10] calculates an ADA (Americans for Democratic Action) score for various news outlets across the political spectrum. In Groseclose and Milyo’s research, ADA is coterminous with political slant or liberal affiliation. The ADA score is calculated from the number of times the news media outlet cites various think-tanks in articles, compared against the number of times Congress cites the same think-tanks in Congressional speeches. The citation patterns frequency determines the news media outlet’s political slant. In the study, editorials or traditional opinionated pieces are omitted as these pieces intend to exhibit biased reporting.

## Data Ideology & Sentiment Analysis:

### Data Sources:

Our project aims to measure bias in news media coverage for the candidates in the United

States 2020 Democratic Presidential Primary. Given the abundance of news media organizations, the group assessed the perceived political affiliations of the major outlets and selected a few that could theoretically represent the complete political ideological spectrum. The selected news organizations include: The Washington Post, The New York Times, New York Magazine, The Huffington Post, The Wall Street Journal, and Breitbart News Network.

The Democratic Part’s first presidential debate ahead of the 2020 United States presidential election was held on June 26, 2019 which drew a total of twenty qualified candidates. As the primary race’s qualifying thresholds increase with each debate, there are currently twelve presidential hopefuls left in the race, with only polling above 3.0%. In addition to filtering the news media sources, the team has selected front-running candidates that encompass the spectrum of democratic ideology, ranging from progressive to mainstream. The selected candidates included in this analysis include: Joe Biden, Pete Buttigieg, Kamala Harris, Bernie Sanders, and Elizabeth Warren.

Finally, the for this analysis the group considered articles from the aforementioned news media syndicates and candidates from September through October 2019. Given the importance of the primary to the United States population and consequently the popularity in its coverage, the group believes that a one-month period is material for the bias analysis.

### Data Acquisition:

After selecting the pertinent news article parameters, the group interacted with News API, newsapi.org, which enables user to search and retrieve news articles from a multitude of different sources based of specific user entered criteria. This REST API service interacts with the group’s python script and retrieves articles that contain the candidate’s search terms in the article’s title or url. By provides these specifications in the load script, the group pre-filtered the qualifying articles for this analysis. For this study, there were X articles considered for the candidates identified from filtered the news outlet sources.

### Sentiment Overview:

Political bias in news articles can present itself in many different forms. As mentioned in the introduction, this project focuses on candidate biases created due to word choice. Words can convey a variety of emotions due to the connotations they carry. To programmatically measure bias, the sentiment of an article or its individual components needs to be assessed.

Sentiment analysis leverages natural language processing, NLP, technology to measure the underlying opinion, mood or emotion of the communicator. For this project, the analyzed data can quantify the journalist’s feelings toward certain candidates and can measure the degree of partiality present at the selected news media organizations.

To measure sentiment, the selected algorithm must be trained to understand the relevant grammatical and wording nuances of political articles. The article’s text also needs to be vetted to remove punctuation, stop words, and tokenize the strings. If properly trained and vetted, the sentiment analysis can capture the categorical and quantitative sentiment measures, polarity and degree, respectively. Polarity describes the analyzed text as “positive”, “neutral” or “negative.” Degree is an outputted score which quantifies the article from extremely negative to very positive.

### Sentiment Process:

Intro Sources:

https://journalism.uoregon.edu/news/six-ways-media-influences-elections https://www.businessinsider.com/what-your-preferred-news-outlet-says-about-your-political-ideology-2014-10 (Paw institute)

<https://www.aim.org/on-target-blog/media-bias-in-strategic-word-choice/>