

## **NAACP Media Research (DS) - V2 - Spring 2020**

Team Members: Shubhangi Jain, Sabrina Araujo, Nada Shalash

### **Motivation/background of the project:**

Today the media plays an important role in shaping not only the opinion of the general masses, but it also influences the law and legislation policy created for citizens. Therefore, it becomes very important that mass communication covers every community in an unbiased manner irrespective of their race, gender or social belonging. This project, in conjunction with the Boston chapter of the NAACP, aims to monitor media reports and investigate media coverage of Boston's black community. We are especially interested in coverage across print and digital channels. We want to focus on what kind of stories are involved with Boston's black community and whether they are positive or negative. Additionally, this study seeks to gauge whether there is a difference in journalism between reports of the black community and other communities. Furthermore, this study wants to decipher the factors that influenced the changes, or more accurately, the evolution of news reports of the black community observed over the past year. As the print and digital media are one of the key channels which provide mass audiences information, we aim to discover whether the information delivered by these channels is in a trusting and supportive format for the black community.

### **Goals, hypothesis, outcomes:**

Under this project, we aim to cover the sentiments of news across various media channels that report the black community. For this, we expect to refine the definitions of black and white geographical communities and information retrieved from the Boston Herald done by the group last semester. We will do additional data scrapes using Word2Vec in order to extract a more meaningful analysis based on topics and storylines portrayed in media relating to black communities in the Boston Globe, Boston Herald, Boston.com, WGBH and WBUR. We will search for people or places recurrently mentioned in such stories. We will search for the explicit mention of race in the media and examine any break of the journalistic convention of omitting race unless relevant to the story, and try to figure out why race was mentioned or if there are differences in the reasons of such mentions between black community members and other people. We are also going to analyze homicide news in recent years in Boston and check the difference in sentiments when reporting the black community in comparison to the white community. In order to carry out some of this analysis, we plan to try to figure out the locations from which news stories originate, and use latitude and longitude data from the US census to learn more

about the demographics of the location of the original story (GIS will be used for this). Finally, the ultimate goal of this study is to create a narrative regarding differences in media coverage of black individuals/communities in Boston in order to assess the possibility of discrimination or variation in treatment.

In terms of our hypotheses of the project's outcome, we expect to find a different narrative between the stories portraying white communities and/or individuals compared to black ones. For instance, we expect portrayal of blacks to lean towards a more negative depiction, such as portraying them as less socially and economically stable.

#### **Non-goals, out of scope topics:**

As part of the project, we aimed to categorize geographic areas based on the business owner in a particular area. This data was defined by the general information of the owners of various companies. There are 2 problems with regards to this:

1. There is a dataset available for minority-owned businesses (blacks, Hispanics, Latinos, women, etc) but it does not specify which business is specifically owned by the black community.
2. Even though there is a more exhaustive database of black-owned businesses available with the local authorities, due to several constraints, our access to the dataset is denied.

As mentioned during the Spark pitches, simply looking at data about black-owned businesses without having anything else to compare it to will not result in any meaningful analysis or narratives about media coverage of the black community in the greater Boston area. For this reason, comparing coverage of black-owned businesses to coverage of other businesses might not be possible unless more data becomes available through Spark.

As our goal is to produce results that show some sort of comparison or highlight differences found in the coverage of the black community, so any other ideas where there is no clear comparison or narrative that can be crafted based on the analysis of the data will not be explored in our project.

#### **End result/product, definition of done:**

This study hopes to provide a storyline of the black community in Boston. This will be done through a comparative analysis of the media portrayal of black citizens to that of white citizens that can be used to improve the portrayal of black individuals in the media. This will build on the work done last semester. The students worked on algorithm design and data scraping of the websites of the Boston Globe, Boston Herald, WGBH and WBUR from 2014 to 2018 using tf-idf. They emphasized on the

portrayal of black communities, defined by tract-level census data. Our analysis will use Word2Vec to extract more substantive words, add an additional scrape of the Boston Herald, a deeper exploration of sentiment and topics in the narratives, and will also look at homicides, specific names of people, and review the explicit mention of race to add on to the work that was completed by previous groups.

### **Open questions, uncertainties**

Differentiating between predominantly-white and predominantly-black communities will prove to be a challenge since there are neighborhoods where distinctions appear at the sub-neighborhood level.

We need to go beyond the words and find topics, or more accurately, narratives, which will require very fine filtering of meaningless words from significant words.