

Elizabeth James, Harrison Richmond, Alex Lin  
CS 506, Lance Galletti  
Project Analysis

## **NAACP Media Research project with BU Spark!**

### **Summary:**

The broader goal of the NAACP is “to ensure the political, educational, social, and economic equality and rights of all persons and to eliminate racial hatred and racial discrimination.” In this project, we worked to analyze The Boston Herald and their coverage of Black communities in order to assess any differences in what stories are told about Black communities. Hopefully, the results will lead us to possible ways in which The Boston Herald can do its part to encourage elimination of racial hatred and discrimination.

### **Methodology:**

#### **Data Collection:**

Our data collection was done through Scrapy utilizing the Wayback Machine. We pulled articles from the local news section of The Boston Herald, aiming to get coverage of the past five years of articles. The Wayback Machine is a log of websites as they existed over time. Scrapy is a tool that allows us to look at the code of the website at these different moments in time to see what was on the web at that moment. There were some difficulties with our collection: The Boston Herald is an interesting site to navigate, as it is constantly changing. The pulls we did ranged over three different URLs for starting the pathway of local news coverage, and four different possible xpath responses from the sites. While all of these pathways were used, there are still chunks of time within the past five years that must have used a completely separate URL or pathway than any of these, as not every time period has coverage. This was not anticipated by us as many other websites, such as The Boston Globe, have only one URL and pathway to follow to cover the entirety of the last five years.

As a result of this, we collected a total of 8,125 articles in the local news section Boston Herald between 2014 and 2019. Of these, only 2,295 articles were found to be classified into a

neighborhood and only 948 were classified as covering a Black neighborhood. We defined Black neighborhoods as those where more than 25% of the population is Black. In general, Black neighborhoods had more coverage than neighborhoods where less than 25% of the population was Black.

**Data Analysis:**

Once we had this data in place, we used python to read in csv files containing the information, and using a tf-idf vectorizer, we were able to compile a list of the top distinct words in each region. Tf-idf is short for a term frequency–inverse document frequency vectorizer, which means that you are exploring the importance of a term to a document in relation to other terms in that document. Using the vectorizer allows these terms to be analyzed numerically.

We used this methodology as we decided it was the best way to gain unbiased information about what the articles were discussing. At first, we had considered compiling a keyword list to search for. We looked at different journal articles of past studies of racism in media and tried to do a preliminary search for the terms these articles had used. However, this brought about another layer of confusion, as we worried it would lead to self-leading analysis. Therefore, we decided to instead use tf-idf to provide the most unbiased view of the data.

**Findings and Observations:**

**Disclaimer:**

Data limitations: This should not be considered a comprehensive list of all articles due to difficulty scraping the site, as described in our methodology.

**Visual Figures:**

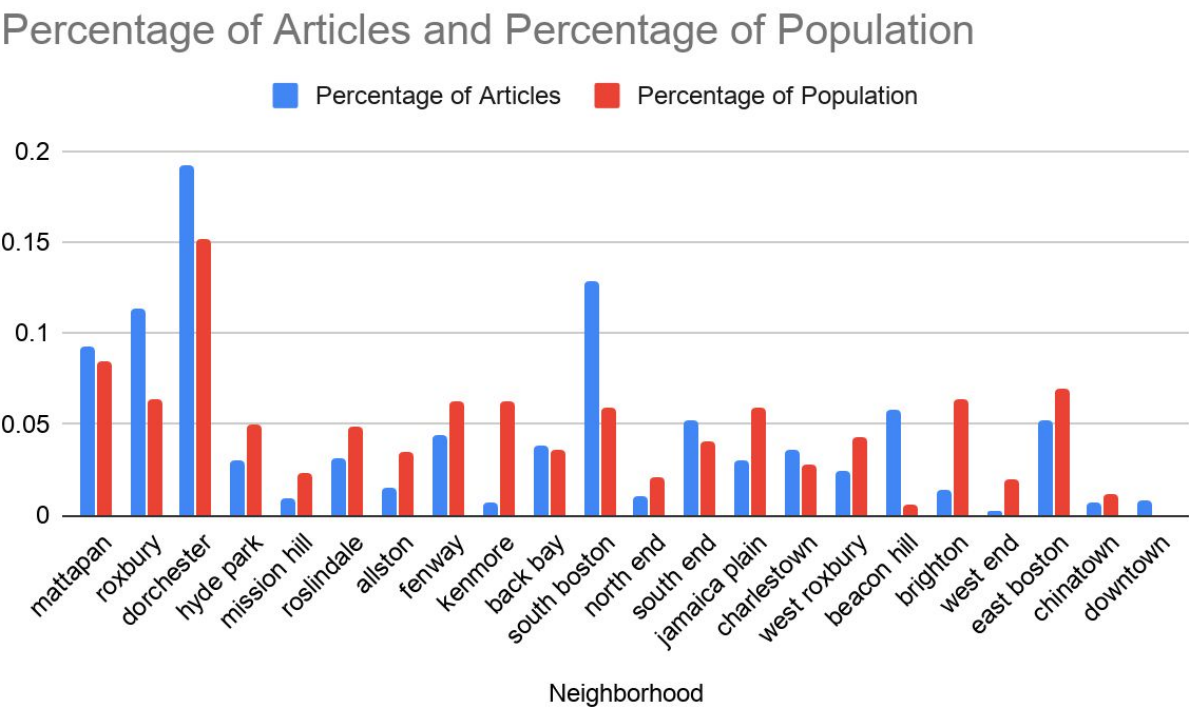


Figure 1: Above is a graph depicting the differences in the percentage of articles we found to be covering a neighborhood in blue, versus the percentage of the population of Boston that neighborhood makes up in red.

Area of Boston	Zipcode(s)	Black Population by Zipcode	Tf-idf: Most distinct words
all studied	all studied	25%	['westin' 'wait' 'think' 'people' 'lot' 'surprising' 'future' 'distant' 'won' 'adding' 'said' 'trump' 'longer' 'little' 'just' 'll' 'said' 'office' 'oval' 'run' 'make'

			'revealing' 'planned' 'time' 'specific' 'chabot' 'hillary' 'cashman' 'jaclyn' 'hosts']
mattapan	02126, 02124	86%, 64%	[ 'contributors' 'restructuring' 'reduction' 'directors' 'area' 'agency' 'half' 'cut' 'additionally' 'percent' 'workforce' 'statewide' 'percent' 'workers' 'social' '124' 'management' 'created' 'additional' 'demands' 'pressure' 'managerial' 'staff' 'supervisors' 'social' 'workers' 'look' 'help' 'report' 'reads']
roxbury	02119, 02121	58%, 69%	[ 'said' 'case' 'officers' 'officers' '11' 'score' '87' 'people' '11' 'joseph' 'degree' 'said' 'walsh' 'simple' 'reviewing' 'exam' 'service' 'civil' 'color' 'walsh' 'said' 'noting' 'officers' 'qualified' 'capable' 'promoting' 'minorities' 'whites' 'cluster' '87']
dorchester	02122, 02125, 02121, 02124	31%, 33%, 69%, 64%	[ 'nice' 'stadium' 'elephant' 'room' 'favor' 'saying' 'going' 'build' 'hard' 'don' 'believe' 'think' 'neighborhoods' 'south' 'end' 'roxbury' 'boston' 'south' 'square' 'andrew' 'stadium' 'olympic' 'potential' 'called' 'lynch' 'place' 'people' 'smart' 'structure' 'know']
hyde park	02136	45%	[ 'gallery' 'including' 'birthday' '23rd' 'amy' 'gave' 'husband' 'necklace' 'jewelry' 'wing' 'daughter' 'pieces' 'identify' 'asked' 'lord' 'gone' 'angel' 'necklace' 'know' '10' 'sneakers' 'sweats' 'dressed' 'u2028alemany' 'away' 'feet' 'sat' 'said' 'mother' 'distraught']
mission hill	02120	25%	[ 'night' 'state' 'current' 'appointed' 'adviser' 'policy' 'senior' 'manager' 'project' 'senior' 'assistant' 'project' '2006' 'intern' 'started' 'lopez' 'lined' 'job' 'does' 'said' 'florida' 'native' 'wife' 'new' 'moving' 'said' 'lopez' 'said' 'issues' 'figure']
roslindale	02131	25%	[ 'games' 'like' 'radicalized' 've' 'peaceful' 'muslims' 'religion' 'cloud' 'cast' 'isis' 'groups' 'hooper' 'terrorist' 'radical' 'said' 'police' 'public' 'safety' 'concerned' 'organization' 'mosque' 'said' 'muslims' 'department' 'rahim' 'plot' 'alleged' 'related' 'crimes' 'hate']
south end	02118	16%	[ 'contributors' 'discussing' 'district' 'nightclub' 'july' '16' '2012' 'ambushed' 'gunfire' 'gray' 'silver' 'suv' 'rhode' 'island' 'plates' 'hernandez' 'held' 'bail' 'murder' 'odin' 'lloyd' '27' 'dorchester' 'theatre' 'leaving' 'dorchester' 'police' 'murder' 'incarcerated' 'hartford']
jamaica plain	2130	13%	[ 'meth' 'gang' '16' 'sentenced' 'ransom' '100000' 'demanded' 'street' 'boston' 'man' 'kidnapped' 'man' 'prison' 'second' 'boston' 'litigation' 'ongoing' 'comment' 'city' 'said' 'walsh' 'martin' 'years' 'julio' 'spokeswoman' 'conspiracy' 'boston' 'victim' 'kidnapped' 'people']
charlestown	2129	9%	[ 'melrose' 'distressing' 'courthouses' 'efficient' 'modern' 'examples' 'good' 'report' 'news' 'alarming' 'committee' 'bonding' 'house' 'told' 'spence' 'month' 'constructed' 'renovated' 'repaired' 'courts' 'decide' 'baker' 'charlie' 'gov' 'consolidations' 'massachusetts' 'majority' 'said' 'courthouses' 'association']
allston	02163, 02134	9%, 6%	[ 'morning' 'told' 'dropping' 'knees' 'prayer' 'cashier' 'begged' '911' 'meng' 'jury' 'shut' 'just' 'said' '911' 'meng' 'later' 'told' 'police' 'finally' 'door' 'equipped' 'mobil' 'said' 'tamerlan' 'shouting' 'expletive' 'surveillance' 'video' 'inside' 'seen']
fenway	01125, 02215	8%, 4%	[ '10' 'upgraded' 'red' 'minds' 'forefront' 'ones' 'loved' 'tonya' 'hospital' 'released' 'statement' 'monday' 'said' 'family' 'condition' 'fair' 'condition' 'said' 'injuries' 'described' 'police' 'center' 'medical' 'deaconess' 'israel' 'beth' 'rushed' 'stands' 'treated' 'win']

kenmore	02115, 02215	8%, 4%	['gallery' 'fenway' 'milton' 'said' 'park' 'people' 'regular' 'prices' 'places' 'reasonable' 'don' 'park' 'hard' 'issue' 'real' 'address' 'idea' 'bad' 'proposal' 'zakim' 'called' 'bouncer' 'local' '29' 'work' 'park' 'just' 'life' 'limped' 'mbta']
back bay	02116, 02199	6%, 0%	['morning' 'sunday' 'said' 'wrote' 'ask' 'letter' 'verify' 'wanted' 'house' 'grow' 'dispensary' 'proposed' 'neutrality' 'support' 'town' 'letters' 'authenticity' 'confirm' 'days' 'past' 'health' 'public' 'department' 'calls' 'taken' 've' 'night' 'richard' 'white' 'town']
south boston	02210, 02127	5%, 7%	['nice' 'stranger' 'expletive' 'god' 'oh' 'said' 'recalled' 'head' 'hand' 'chase' 'gave' 'alemany' 'got' 'alemany' 'situation' 'ran' 'break' 'saw' 'follow' 'ordered' 'wall' 'threw' 'shirt' 'picked' 'looking' 'alemany' 'said' 'told' 'hours' 'body']
west roxbury	02132	6%	['said' 'meg' 'options' 'charter' 'private' 'chosen' 'families' 'boston' 'engage' 'boston' 'families' 'new' 'entice' 'families' 'boston' 'wants' 'face' 'need' 'voice' 'school' 'current' 'claudio' 'neill' 'michael' 'chairman' 'committeemen' 'school' 'join' 'committee' '2018']
beacon hill	02108	5%	['families' 'important' 'hill' 'institute' 'said' 'salaries' 'boston' '2024' 'staffers' 'public' 'transparency' 'include' 'tuerck' 'disclosure' 'salaries' 'working' 'behalf' 'organization' 'sponsoring' 'bringing' 'olympics' 'tuerck' 'beacon' 'david' 'particularly' 'organizing' 'registered' 'organization' 'comply' 'statutory']
brighton	02135	5%	['morning' 'good' 'tamerlan' 'praised' 'told' 'good' 'job' 'boy' 'job' 'complied' 'parked' 'pump' 'shell' 'gas' 'station' 'memorial' 'said' 'said' 'eliminate' 'answer' 'meng' 'roommate' 'trying' 'reach' 'cellphone' 'said' 'phone' 'meng' 'say' 'single']
west end	02114	5%	['report' 'community' 'living' 'stories' 'shared' 'white' 'father' 'help' 'came' 'just' 'recovery' 'horne' 'people' 'lot' 'look' 'parish' 'paul' 'vincent' 'st' 'member' 'southie' 'growing' 'projects' 'development' 'come' 'll' 'hoping' 'later' 'ya' 'goodbye']
north end	02110, 02019, 02113	4%, 4%, 0.7%	['end' 'cold' 'line' 'stood' 'stars' 'news' 'broadcast' 'boston' 'list' 'long' 'including' 'mourners' 'wher' 'mattapan' 'year' 'past' 'calls' 'multiple' 'responded' 'police' 'boston' 'cared' 'people' 'wish' 'just' 'jacobson' 'natalie' 'hours' 'yesterday' 'said']
east boston	02128	3%	['plans' 'laughed' 'answer' 'charged' 'open' 'door' 'bus' 'overwhelmed' 'got' 'stomped' 'head' 'don' 'remember' 'man' 'said' 'revere' 'man' 'refused' 'smith' 'feet' 'said' 'throw' 'snowball' 'driver' 'alleged' 'victim' 'said' 'doesn' 'recall' 'throwing']
chinatown	02111	2%	['wearing' '100000' 'dorchester' 'cases' 'assault' 'open' 'citing' 'bail' 'cash' 'held' 'new' 'downing' 'recommended' 'galatis' 'act' 'intentional' 'accident' 'joun' 'district' 'court' 'joun' 'revoked' 'attack' 'intersection' 'standing' 'hearing' 'pretrial' 'u2028aug' 'pending' 'cash']
downtown	02203	0%	['calls' 'yesterday' 'police' 'told' 'woman' 'said' 'police' 'transit' 'morning' 'said' 'walked' 'police' 'court' 'municipal' 'boston' 'today' 'arraigned' 'expected' 'assaulted' 'orange' 'expletive' 'police' 'ready' 'ready' 'said' 'arrested' 'told' 'estrada' 'approached' 'away']

Figure 2: Above is a chart depicting the neighborhood of Boston, the Zipcodes that make up that Area of Boston, the Black Population percentage for that neighborhood, and a list of the top unique words discovered by our tf-idf analysis. Rows in blue are neighborhoods we classified as Black Neighborhoods.

### **Analysis:**

The information above is very telling about how The Boston Herald handles communities of color as opposed to other communities.

The first event of note is the discrepancy between the percentage of articles we found to be covering Black neighborhoods versus the percentage of the population of Boston that those neighborhoods made up. As we can see in Figure 1, the top three neighborhoods- ranked by the percentage of their inhabitants that are members of the Black community -all compromised a larger portion of news coverage than they did of the greater Bostonian population. While the fourth neighborhood categorized as a Black neighborhood did not follow this pattern, it is still of value as out of the eighteen neighborhoods that were not categorized as Black neighborhoods, only four compromised a larger portion of news coverage than they did of the greater Bostonian population. This means that on average, Black neighborhoods were more commonly covered by the Boston Herald than other neighborhoods in relation to the percentage of the city's population they make up. We had hoped to find out more about the volume of media discussing Black people and communities, and here we find that it is even more coverage than one might expect.

To gain some more insight into what these media reports are saying, we can look at the words in Figure 2. To focus our study, we have summarized the general content associated with these words for all of the communities classified as Black neighborhoods to aid in our discussion of the topics that were common:

Mattapan: Workers & Unrest

ex: workforce, management, staff, demands, pressure, help

Roxbury: Officers & Racial Discussions

ex: officers, case, civil, color, white, minorities

Dorchester: Places & Events

ex: neighborhoods, end, square, stadium, build, olympic

Hyde Park: People & Emotions

ex: daughter, husband, mother, gone, distraught

What is important to note here is that, while all of these neighborhoods were classified as Black neighborhoods, there was still a vast difference in the topics that were most commonly discussed. From the start of our project, something we wanted to discuss was what topics are being covered in relation to Black people and communities? Our original prediction was that Black neighborhoods were less likely to be covered in feature and community stories of triumph or heartbreak, and more likely to be covered in news reports and current events. Our results show that the answer is different for different black communities.

Roxbury is most similar to what we would've expected to see: discussion of officers often is tied to discussion of police activity, and the word 'case' is often used in relation to judicial happenings. What was additionally incredibly intriguing about this particular neighborhood's results is that there were results that suggest lots of discussion surrounding race, using words such as 'color', 'white', and 'minorities'. It is difficult at this time to say in what context race was being discussed- if the articles were being aware or racist is unknown to us- however the fact that it is in discussion is still incredibly important.

Mattapan is the next most similar to what we would've expected to see, as our results depict that the main topics of discussion were in relation to workers and unrest going on in the community at that time. These are still current events occurring within the community, even though they are more ongoing than what we had expected, as our expectations were more in relation to fires, robberies, winter storms, interpersonal violence, and the like. However, we did expect for the topics to be more negative in nature, which this unrest does fall into.

The data for Dorchester, on the other hand, still points to discussion of current happenings, but more positive ones than we had originally expected. Discussions of the Olympics and building. Common words also included 'smart', 'know', and 'believe', which are all often used in positive contexts surrounding community happenings.

Hyde Park is the furthest from our expectations, as data suggests most of the topics in this area were community stories, as key words popped up that described family members-

‘daughter’, ‘husband’, ‘mother’, events- ‘birthday’, and gifts- ‘jewelry’, ‘angel’, ‘necklace’. These highly suggest feature articles about community members and their experiences.

When looking at the other neighborhoods, we see other words relating to current events - ‘victim’, ‘police’, ‘responded’, ‘assault’ - as well as other words relating to community stories - ‘recovery’, ‘families’, ‘memorial’, ‘school’. This leads us to believe that while there are some areas of difference in how The Boston Herald reports on Black neighborhoods, it is not as extreme of a difference as we originally imagined.

### **Conclusion:**

What we have seen through our observations of The Boston Herald is that while they do report upon Black communities more often, this does not necessarily correspond to a change in the overall topics discussed. Our data suggests that articles about Black communities still contained feature stories, current events, politics, and more, just as did articles about other communities. This was exciting to see as our original prediction had been that there would be extreme negative bias in how these neighborhoods were represented, however our data does not show that.

Given, this does not mean that these biases do not exist. Although our data does not show it in an obvious manner, it does not mean that there aren’t differences in the ways in which these stories are being talked about in terms of their tone and sentiment. Additionally, our data pull is not representative of all of the data that is out there.

Overall, it is important for news outlets to highlight both current events as well as community ongoings for all neighborhoods equally in order to inform the general public of the opportunities and happenings within these different neighborhoods.