

Introduction

- 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.

Methods

Boston Herald:

- *Boston Herald* – A news website that reports upon current events, natural disasters, community stories, and more. It has been around since 1846, though we focused on the past 5 years.

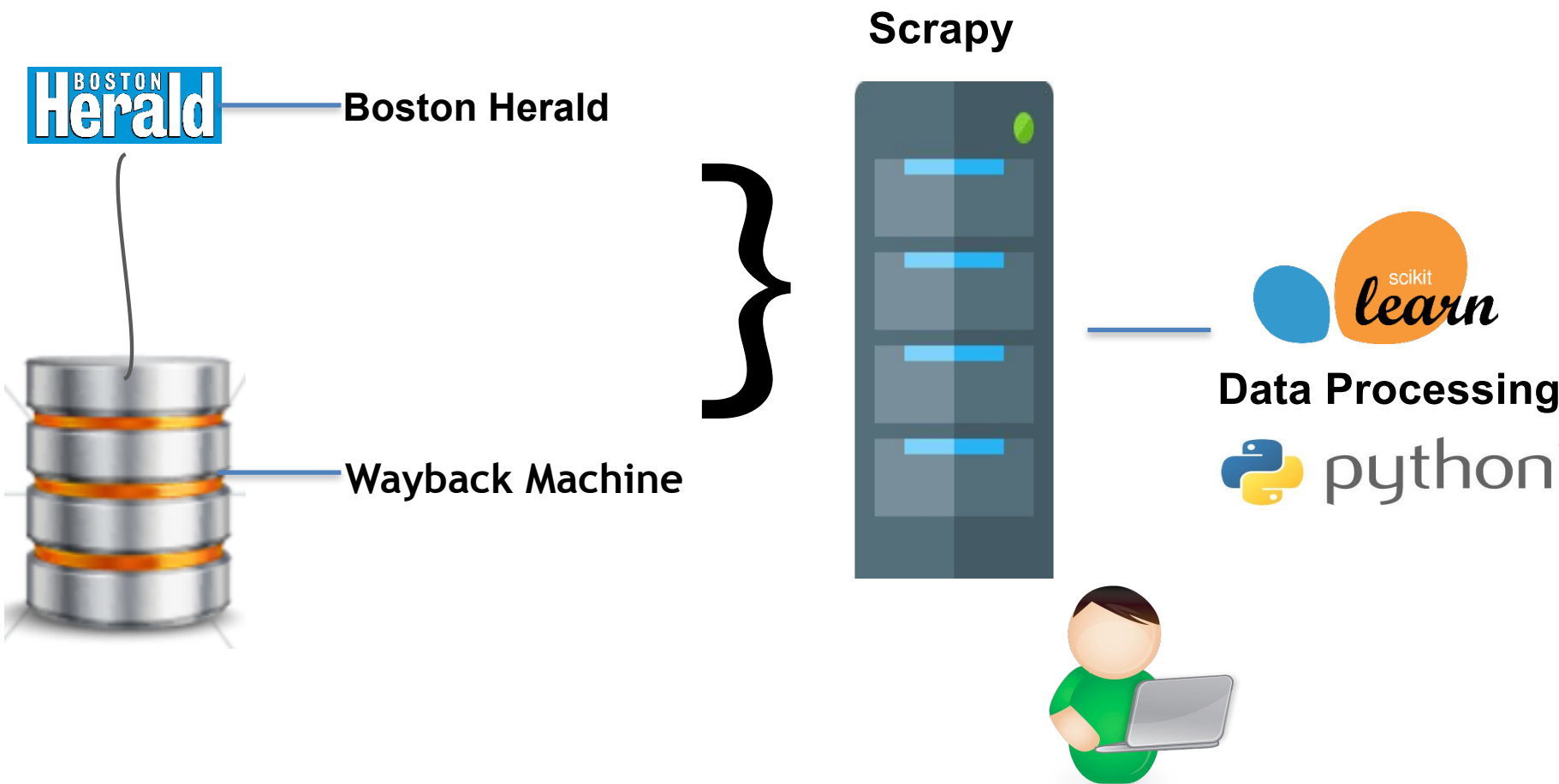
Python Scrapy & Wayback Machine:

- *Wayback Machine* - a log of websites as they existed over time; allowed us to view the local news page of the Boston Herald website as it existed from 2014-2019.
- *Scrapy* – Python package that aids in web scrapping; uses xpath methods to scrape articles’ text.

Term frequency–inverse document frequency(Tf-Idf):

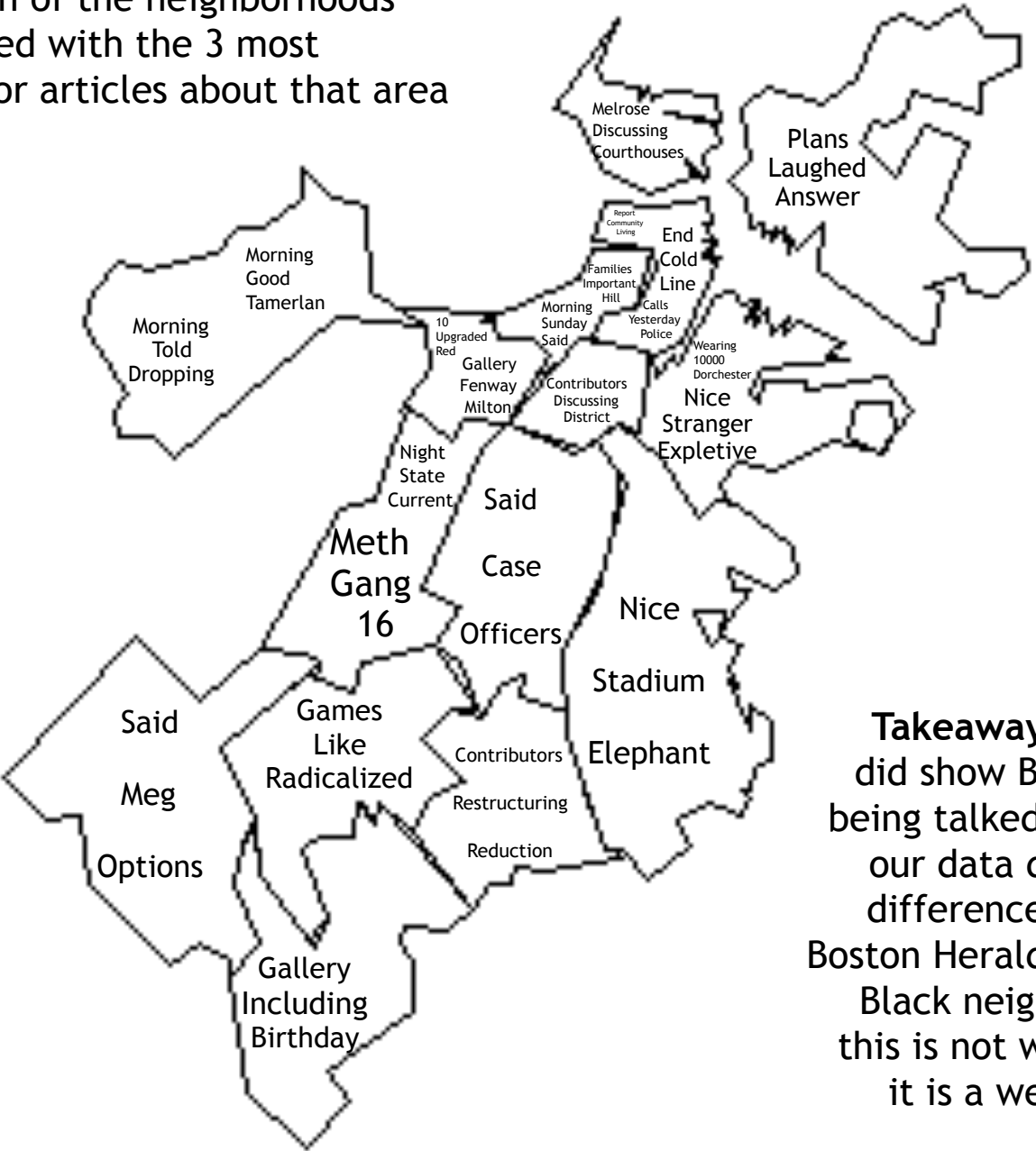
- *Tf-idf*: allows exploration of the importance of a term to a document by showing its value in relation to other terms in that document. It helps remove biases in the data by making this comparison.

Visualization



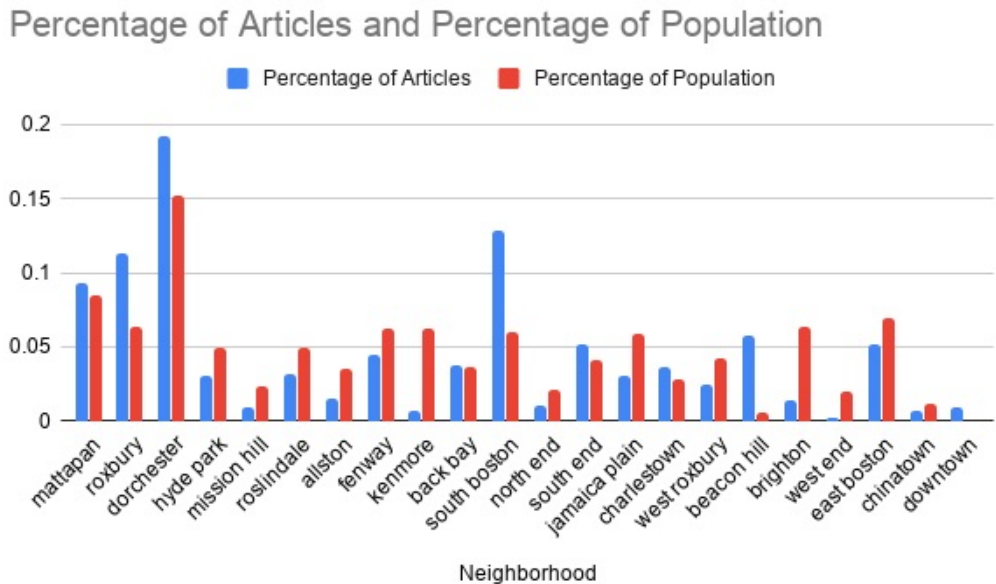
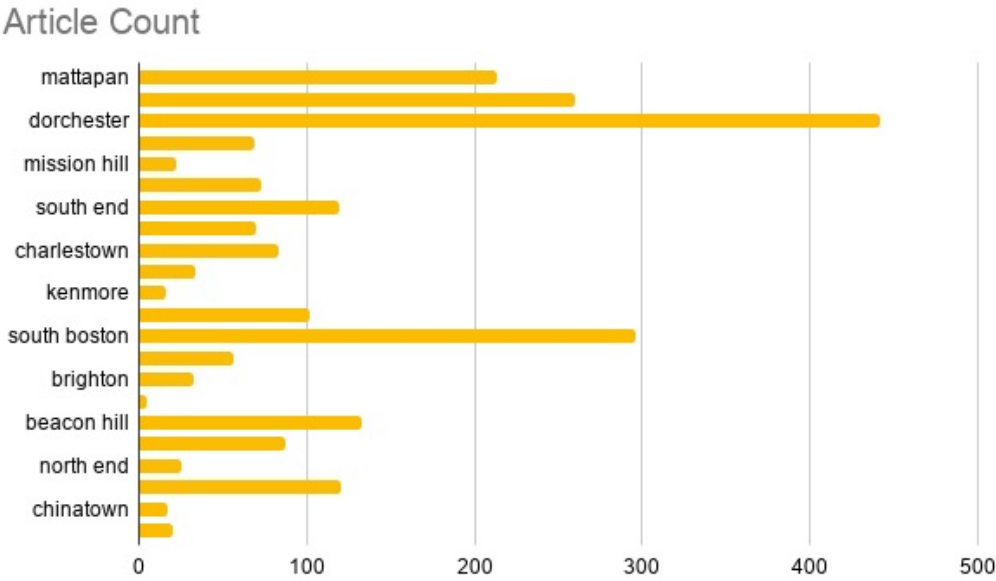
Results

Map: Shows each of the neighborhoods of Boston labelled with the 3 most distinct words for articles about that area



Takeaway: While our data did show Black communities being talked about more often, our data did not show vast differences in the way The Boston Herald was reporting upon Black neighborhoods. While this is not what we predicted, it is a welcome surprise.

Data



Area of Boston	Zipcode(s)	Black Population by Zipcode	Tf-idf: Most distinct words
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' 'away' 'feet' 'sat' 'said' 'mother' 'distracted']