New England Blacks in Philanthropy

Financial Support Given to Black Non-Profits

Shunying Chen (U47342767) | Elsa Jaysing (U77504656) | Reha Patel (U17557726)

Current Pipeline

1. *Find Grantees*
    a. Download 990 Form from Guidestar
    b. Convert pdf to excel
    c. Extract EIN, Amount of Cash Grant, and Purpose of Grant or Assistance (Columns B, D, H)
        i. Can use EIN to search up organization in Guidestar to obtain name if needed
2. *Obtain financial information and board of directors names*
    a. Use EIN obtained above with the Guidestar API
        i. Returns financial information and names of board members
    b. Convert JSON to CSV File
        i. Use converter to temporarily display what data will look like: https://json-csv.com/
3. *Demographic Information*
    a. Search/lookup name and company using RocketReach API
    b. Get LinkedIn profile url
    c. Using a LinkedIn scraping tool, find the person with the url and save results
        i. Returns name, about, experience, education, interest, and accomplishments

Comments on Pipeline

1. *Find Grantees*

    The caveat with this approach is that you have to manually look up and download the 990 Forms from Guidestar to get the Grantee information as the Guidestar API doesn't return it in JSON. We are still looking for ways to get around this so that this can be done in bulk.

2. *Obtain financial information and board of directors names*

    For now, we've been using a converter to see what the data looks like to make sure we're able to get the information we need. We've been trying to figure out how to do it with code; however, since the JSON is very jumbled we haven't been able to use the standard

approach to convert it. This step is something we're still trying to perfect so that when the code is run it neatly puts everything into a CSV file. A potential alternative to this that we are currently exploring is the Charity Navigator API. Rather than returning the entire 990 as Guidestar does, Charity Navigator may potentially only return the fields needed for this project: key financial data, board of directors/executives, and the board salaries.

3. *Demographics*

   The main use the RocketReach API has is obtaining the LinkedIn profile url, since the scraping we're currently using needs the url to scrape. The caveats with using the RocketReach API is the following:

   - Lookups are limited to five per API key so you'd need to buy a subscription
   - There is a problem with the name and company searches as the names provided from the 990 Form don't always match up (nicknames, using middle names, etc.)
   - Multiple results sometimes show up even when using both name and company
   - No results sometimes show up when using both name and company, but results show when using just the name

   Because of these complications with the RocketReach API we've decided to find a way to skip that step since the LinkedIn scraping seems to be a good way to obtain further information on the individuals.

Next Steps

We're looking into using another API called Charity Navigator to see if we can obtain the same information we're getting from Guidestar. While the Guidestar approach works, we still want to obtain more data for our client, and due to the limitations of lookups we have for Guidestar, we already know we would only be able to get information for only one Community Foundation, The Boston Foundation. In addition to there being a limited number of lookups associated with Guidestar, the API also has a throttling limit which would make the process of looking up charities take days even if run constantly.

As mentioned above we're still working on how to code converting the JSON from Guidestar to a CSV file and once we know what that looks like we'll preprocess the data so that it only displays the financial information and board of directors names. However, because the JSON

from Charity Navigator has fewer nests, it is also possible that using it will resolve our problems of converting JSON to CSV.

Since we need the LinkedIn url for the LinkedIn scraping approach we've taken, we're looking into other ways to scrape with just the name and not the url. Once we have that down we'll parse the scraped results and cross-reference with the keywords list provided by our client to check if there's any associations with Black communities. Currently, we are trying to build and customize our own LinkedIn web scraper from related resources found online.

Answering Questions

Since we haven't solidified a good way to obtain association information for each individual on the board of directors, any questions we would be answering would be related to financial information. Since we're only able to collect data for one community foundation - The Boston Foundation - we don't feel that it's appropriate to answer any of the questions our client had given the fact that it would be about only one foundation out of many. We think that the questions our client has are very serious and important questions to answer, which is the main reason we were so thrilled to be working on this project, but to that same sentiment we don't want to draw false conclusions simply because we have some data and want to present some type of answer.

Connecting Project to Next Semester

Since it's evident that this project would need the help of at least one more semester to actually collect all of the data past just the Boston Foundation, we're still focusing on solidifying a proof of concept and then coming up with ways to analyze the data once it's obtained to then more easily answer those questions. We will hopefully have not only a pipeline for collecting the data, which we already have a good chunk of, but also for data visualization and other ways to help analyze areas collected like the financial data. This will be represented in a way similar to what we have above so that whoever is able to obtain all of the data won't have to start from scratch when trying to answer the questions provided.