NP DataHub

Repository:
https://github.com/babyspider/NP_DataHub/
This document in markdown:
https://github.com/babyspider/NP_DataHub/blob/main/Docs/Project_Proposal.md

Overview

NP DataHub will be a website for other non-profits, policymakers, and donors to view information on non-profits. This will help the end-user in making informed decisions. We will pull from the publicly available data on nonprofits produced by the IRS. Then through deploying machine learning algorithms, we will gather an in-depth analysis of all non-profits we have data for. Finally, we will publish our findings on a pretty website.

Semester Plan

While our end goal for the project is an all-encompassing website. We will start off small choosing only one NTEE code to go in-depth on. Following choosing an NTEE code, we will begin gathering all relevant data and storing everything in a database. Once we get a database set up we can start implementing some basic algorithms just so we can get some output to put up on our website. We would then like to explore some interesting algorithms we can design for fiscal variables and maybe even be able to measure the impact of organizations within an NTEE sector.

Technology

We plan on using an SQL Database hosted on AWS or Azure to store all of our data from the IRS. Then we will use Python to analyze the data. From there we will turn our findings into graphs or diagrams using a python library such as Plotly or other Open Source data visualization tools such as these. Finally, we will publish our findings on a website using Javascript.

Group Members

Name	Github Handle	Email
Joey Hinckley	JoeyHinckley34	hinckj@rpi.edu
Ariana Lissak	babyspider	lissaa@rpi.edu

MileStones

- Week 1: Identify one of 26 NTEE sectors within New York State, which will serve as a testing sample for algorithmic construction, platform development, and data visualization implementation.
- Week 2: Establish the preliminary database foundation for the NTEE sample and associated functionality to be implemented within the V1.0 platform.
- Week 3: Establish the V1.0 platform with associated end-user functionality for data visualization/dashboard construction and algorithmic results.
- Week 4: Identify one public database source that aligns with the chosen NTEE sector to be included in advanced algorithmic construction. (Basically, what are we trying to tell the end-user or why are these results important?)
- Week 5: Begin testing phase on selected fiscal variables of individual nonprofits and associated NTEE codes and define their purpose. (What story are we trying to tell through data that will serve the public interest based on fiscal performance?) Developers can establish the number of algorithms with a focus on quality.
- Week 6: Begin testing phase on selected fiscal variables of individual nonprofits and associated NTEE codes and define their purpose and alignment with the chosen public dataset. (What story are we trying to tell through data that will serve the public interest?) Developers can establish the number of algorithms with a focus on quality.
- Week 7: Finalize algorithms with an emphasis on accuracy and precision.
- Week 8:Incorporate findings within the chosen open-source data visualization libraries.
- Week 9: Construct a data visualization dashboard.
- Week 10: Launch V1.0 of the platform for user testing and presentation.