# Design Document

# Objective

Our goals for this project are to build a web scraper that searches business pages on CNN from mid-late March of 2020 until April of 2020 to see the amount of chatter created due to the CoronaVirus.

# Background

Time period -Covid ( March 2020 - April 2020) High unemployment due to multiple business closing due to Governor's Mandates -Unemployment became more valuable than people’s regular jobs so a lot of people quit their jobs PPP Loans were a huge topic around this time because while a lot of small businesses High inflation -value of the US dollar decreasing A lot of talks was around the country’s debt and ability to survive Covid -Sustaining hospitals -Providing stimulus checks for financial relief

# Data Collection

Describe your preliminary ideas for data collection and cleaning. We will primarily use the CNN dataset containing over three hundred articles. We will analyze the data and manipulate it to see which words were used the most within a certain period of time.

# Detail Design/Ideas

Use Natural Language Processing to interlink computer and human language, allowing users to analyze a large set of data to create the web scraper.

# Project Plan

-Presentation -Prototype -Code Intermediate Goals -Analyze data, interpret the findings- -Pseudocode Overall Things to Do -Analyze Data-Create Model-Trim Model-Rehearse presentation

# Fairness Considerations

Complete the ethical storyboarding activity. (completed) There are largely positive implications to completing our model, however we must be wary of the ways our model could negatively impact someone as well as the potential biases of the data.

# References

Provide an annotated list of reference materials that you used in preparing these design documents, as well as any additional references that you plan to use in your project.

* MatplotLib
* Numpy
* Seaborn
* Pandas
* Scikit Learn