**Introduction & Context**

* Since the pandemic began multiple programs have been aimed at alleviating the financial burden from COVID, most notably was the Paycheck Protection Program (PPP).
* Two facts informed our interest in studying this program. First, statistical machine learning (SML) algorithms were used to automate the loan approval process. And second, there have been allegations of fraud upwards of $80 Billion USD.
* We are interested in evaluating whether key factors – like business type or amount forgiven – can be predicted using ML with the data that lenders themselves had access to.
* Our final dataset was collected from the Small Business Administration (SBA) and includes data for all PPP loans disbursed for businesses located in the District of Columbia.

**Approach**

* Our goals are twofold. First, can we classify the type of business (e.g., non-profit, profit, etc.) an applicant is based on loan data. Second, can we predict the amount forgiven based on descriptive data (e.g., number of employees, industry, etc.).
* To accomplish these tasks, we use some of the following methods: Logistic, LASSO, Linear, Ridge Regression; Linear Discriminant Analysis; Quadratic Discriminant Analysis; Random Forest.
* We find that the best results for classifying based on business type came from a [METHOD HERE: KNN, LDA] model. While the best results for predicting amount forgiven were from a model built using [METHOD HERE].

**Assumptions & Limitations**

* We assume that the data collected by the SBA is accurate for all loans.
* Additionally, we are limited by data availability. We recognize that other important variables/factors are likely not included in our dataset.
* Our results are not generalizable beyond the District of Columbia.