**Neural Network Report**

**Analysis Overview**

* Using a large .csv file and pandas, the data was manipulated in order to better fit into a neural network model used to predict effectiveness of funding that was provided by Alphabet Soup.

**Results**

**Data Preprocessing**

* The variable used as the target was the “IS\_SUCCESSFUL” column in the dataframe
  + This was a binary indicator that determined whether or not previous projects were successful or not
* Feature Variables
  + APPLICATION\_TYPE
    - The Alphabet Soup application type. This column was trimmed to include the top 5 values and an ‘other’ bin that was a aggregation of all other values for a total of 6 values
  + CLASSIFICATION
    - A code that classifies the government organization. Following the the same path of preprocessing as the APPLICATION\_TYPE
  + ASK\_AMT
    - Indicates the amount of cash asked for. This column was binned into 4 categories 0-10000, 500000-100000 500000-1000000
  + INCOME\_AMT
    - An income classification. This column was already cleaned to include bins of income ranges
* Removed variables
  + STATUS
    - Indicates if the organization is active still. A bit redundant as its assumed that successful orgs are likely still active
  + SPECIAL\_CONSIDERATIONS(Y,N)
    - Indicates if there were any special considerations for the application or not. Removed to simplify dataset

**Compiling, Training and Evaluation**

* The closest the model came to 75% accuracy was on the 3rd optimization attempt
  + A total of 225 neurons were used spread across 4 hidden layers
    - Adding more neurons and layers proved to be moderately more successful in accuracy across the 3 optimization attempts
  + Unfortunately the model did not achieve target performance
  + Adding more layers and neurons along with removing the ‘STATUS’ and ‘SPECIAL\_CONSIDERATIONS(Y,N) columns bumped the accuracy from 72% to 74% accuracy

**Summary**

The deep learning model construction was in parts both basic and rather complex in certain aspects. There weren’t too many variables to account for leaving the preprocessing to be relatively straightforward. The optimization attempts however did not yield the desired target accuracy. The changes that were made in these attempts very minimally affected the performance. I believe that the optimization would be easier and more fruitful if a PCA model was conducted on the dataset. Ordering the features that account for more variability within the target would likely help construct and better performing deep learning model