**Analysis Report on Alphabet Soup Charity Optimization.**

**Overview of the analysis: Explain the purpose of this analysis.**

**Data Preprocessing**

**1. What variable(s) are the target(s) for your model?**

**Ans:** The target for the model in all the three files and models is the column "IS\_SUCESSFUL".

**2. What variable(s) are the features for your model?**

**Ans:**  
**A. Starter\_Code.ipynb :**

There are 43 variables used as the features in the original Starter code. This includes columns Status, Ask\_amnt, ApplicationType\_other, Application\_type\_T10, Application\_type\_T19, Application\_Type\_T3,Application\_Type\_T4, Application\_Type\_T5,Application\_Type\_T6,SPECIAL\_CONSIDERATIONS\_1-9999,SPECIAL\_CONSIDERATIONS\_10000-24999,SPECIAL\_CONSIDERATIONS\_100000-499999,SPECIAL\_CONSIDERATIONS\_10M-50M, SPECIAL\_CONSIDERATIONS\_1M-5M, SPECIAL\_CONSIDERATIONS\_25000-99999, SPECIAL\_CONSIDERATIONS\_50M+, SPECIAL\_CONSIDERATIONS\_5M-10M, INCOME\_AMT\_N ,INCOME\_AMT\_Y

**B. AlphabetSoupCharity.ipynb**

The same 43 variable as used in Starter Code are used as features in the attempt for optimization in the model for the file AlphabetSoupCharity

**C. AlphabetSoupCharity\_Optimization.ipynb**

The last attempt for the optimization was made with some modification to the features. There are 42 features after some modification. This includes ASK\_AMT, APPLICATION\_TYPE\_Other, APPLICATION\_TYPE\_T19, PPLICATION\_TYPE\_T3, APPLICATION\_TYPE\_T4, APPLICATION\_TYPE\_T5,APPLICATION\_TYPE\_T6, AFFILIATION\_CompanySponsored, AFFILIATION\_Family/Parent SPECIAL\_CONSIDERATIONS\_1-9999SPECIAL\_CONSIDERATIONS\_10000-24999, SPECIAL\_CONSIDERATIONS\_100000-499999,SPECIAL\_CONSIDERATIONS\_10M-50M,SPECIAL\_CONSIDERATIONS\_1M-5M, SPECIAL\_CONSIDERATIONS\_25000-99999, SPECIAL\_CONSIDERATIONS\_50M+,SPECIAL\_CONSIDERATIONS\_5M-10M,INCOME\_AMT\_N,INCOME\_AMT\_Y

**D. What variable(s) should be removed from the input data because they are neither targets nor features?**

I have removed the variable 'STATUS' as it does not affect the features or the targets.

**Compiling, Training, and Evaluating the Model**

**How many neurons, layers, and activation functions did you select for your neural network model, and why? Were you able to achieve the target model’s performance?**

**What steps did you take in your attempts to increase model performance?**

The original **Starter\_Code.ipynb** has the basic neural network where there are 43 input variables. There are 2 hidden layers and both have 86 neurons. The general rule of thumb states to have 2-3 times the neurons as the input variable. Also the activation function used was relu in the 1st and 2nd hidden layer and Sigmoid in the output layer with 1 neuron. The total number of epochs are 50. The accuracy for this model is 0.728 which is less than the required optimization.

To get the desired optimization, I have made changes in the model layers and number if neurons and also reduced the input variable.

In the second attempt as in the **AlphabetSoupCharity.ipynb** there are 43 input variables but the 1st layer has 129 neurons and relu activation function, the 2nd hidden layer has 129 neurons as sigmoid activation function, 3rd hidden layer has 86 neurons and sigmoid activation function and the output layer has the 1 neuron and sigmoid activation function. The number of epochs used to train the model is 100. The accuracy for this model is still 0.728 which is like the previous model.

In the next model in the **AlphabetSoupCharity\_Optimization.ipynb** there are 42 input variables where the column of STATUS was additionally dropped. The cutoff for the classification was reduced to 1000. In this model there are 3 hidden layers where 1 layer has 111 neurons and relu activation function, 2nd layer has 74 neurons and relu activation function and 3rd hidden layer has 37 neurons and Sigmoid activation function. The last model is also trained on 100 epochs. Even though there have been some reduction in the input variables and the number of hidden layers and the neurons, the accuracy of the model did not change much. The accuracy remained almost the same at 0.728 which was close to the required model performance if not the same at 0.75.

**Recommendations:**

There is a chance that if original data which was used to study is modified or more factors which have higher influence on the output result are chosen then there could be improvement seen in the model performance. There is also a possibility that if the model is trained on higher number of epochs it can yield better result.