CS 241 Final Project Write Up
Will Compton

To understand the and run a script on a data set we need to know what the data set is. For this project I used the FEC donations data from the 1/1/2021 until 12/12/2021. From that data I can use the script to understand things within it. In the current set up you can look up the name of someone who has donated during the time frame. It does this by sorting the information as a binary search tree(BST). When you ask for every donation with the name it will go through that tree and return every donation the person with that name has contributed in the order of donor, donee, and amount. With some minor changes due to the way the script is set up you can look up all of the donations to a certain donee and you can also look up all of the donations of a certain amount. There is a binary search tree for each of these three data fields so just switching which tree the script is searching through will allow these different results to be captured.

For insert I have it check to see if the BST has anything in it and if not to create a new node for the three important pieces of data. I used a BST because the data gets sorted throughout the BST as it gets inserted. There are three different trees and they are for the Donor, Donee, and Amount. Since each BST is inserted individually all three trees are able to be looked through for data.

Contains has two different functions in the script. The first one is in the main class and just calls the one in the subclass that is able to recurse through the BST and see if the data is inside of it. In order to get all of the times the data is seen in the BST I set up a data field where each time the data is found it is added to the field. Then, at the end all of the times the data is seen it is returned. I chose this method because it is very quick going through the BST with a best case of $O(\log(n))$ and a worst case of $O(n^2)$. It takes a long time to insert the data in the beginning, but very little time to look up a data piece.

As I have stated before I have implemented the sort through using the BST for the overarching structure of the data when I insert it. I sort the data at the very beginning and am able to use that sorted data to easily check to see if the data contains what I am looking for.

From this data I was able to learn that every time I looked up someone their name came up multiple times. This was very interesting to me because it seems that people who donate once are way more likely to donate again. This could also be an issue with my code, but I am not sure where the issue would be. I looked at the zip code where I used to live in Alaska because I was curious to see the diversity in what party was receiving the majority of the political donations. I know that Anchorage is very conservative, but I was curious to see how many more liberal politicians/groups were getting donations. I was able to find that there were very few donations to more liberal leaning politicians/groups. I was expecting this just due to the longstanding tradition of Alaska being a red state.

This project was very satisfying to do when I was able to see all of the data being ingested and then spat back out for the insert and contains functions. The only real annoyance I had was that I was unable to use a larger data set. I tried using another zip code from where I lived in Michigan and the data set was then about 8,000 lines. I would run into an error of max recursion depth reached. Thank you for the majority of the code on this assignment. I am sure I would have been able to make the recursion to make sure all of the times the data is in the set it gets printed.