Problem Definition:

The dataset I analyzed was the Titanic and the spreadsheet showed several parameters that were logged. From age, sex, and even how many family members were together was recorded. What I want to analyze from this dataset is what was the likelihood of someone surviving the tragedy. To find this answer I will be using a decision tree and level of importance to see one’s best-case scenario.

Final Remarks:

After analyzing and cleaning the dataset, I found that when it comes to the level of importance, the passenger’s sex was the most important factor tested. Looking at the raw data without any insight, you would notice that the female passengers had a greater chance of surviving the sinking of the ship. This could be because of several reasons one being, there was a call to action to save the women and children of the ship. Next on the list would be the Pclass meaning where your room was on the ship. If one person’s cabin was closer to the floor of the ship the likelihood of them surviving would be less than someone having their cabin towards the top. Finally, was SibSP or the amount of family one had on the ship. What I’m assuming is that because women and children were the main priority of saving, they would stay together while the men would help in rescue efforts. Of course, this isn’t a direct match for siblings, parents, and spouses to survive together. However, I do believe because that because the sex of the passenger was very important in this analysis, than I conclude that women had a better chance of surviving compared to men.