Figured I would write a short report detailing my experience with the last project and what went wrong, vs how I fixed it in this project and managed to get a more accurate prediction. First, for my last project I noticed that my accuracy was unnaturally high. It was hovering at around 96%. I realized this was because of two reasons: my dataset was biased towards it usually raining vs it not usually raining, and my dataset train test split was not big enough. I fixed both, but the accuracy was still very high, and the dataset did not really make sense. I then realized that my API for some reason was not pulling weather data from Boise, Idaho like I expected but rather from somewhere in Washington called Boise.

After realizing this, I corrected my mistake and acquired a new dataset from the zip code 83706 which behaved more like expected, with most days not having rain. This had a accuracy of around 83% as a baseline, making my new goal at least 88%. I was able to hit this goal easily by adding another year of data to the dataset, as that boosted the accuracy to about 89%. Finally, I removed some of the redundant columns on my dataset and that boosted the accuracy by a further percent or so. After this, I attempted to use scaling to try and scale my pressure to my temperature, as I found that pressure was very important to determining rainfall but was being treated as unimportant due to temperature being such a high range of values. Unfortunately, I was unable to figure out how to scale the two different types of data with each other properly. I tried min max and Standard Scaler. I think standard scaler is the correct one to use in this scenario, I was just unable to get it to work properly. I gave up and figured that since I hit the 5% improvement, I could just write a report explaining my circumstances since I managed to fulfill the assignment requirements.