Tyler Smith **Project Proposal** 5/2/2024

The prior proposal outlined three possible ideas for predictive modeling - a precipitation prediction model, a veterinary dental model for predicting whether certain types of breeds of dogs were prone to periodontal disease, and a predictive model using World Bank economic data to see if the model could classify the regions in the world where the economic data was gathered. The precipitation model was the main model of interest, but after the data wrangling and exploration, I have decided to go a different way. Mainly due to the fact that I need more features to build an adequate precipitation model than I have at my disposal.

My next idea was to use my employer's customer expenditure data to make a membership churn model, but after talking to my boss I decided to use a Bank Customer Churn dataset from Kaggle (https://www.kaggle.com/datasets/radheshyamkollipara/bank-customer-churn), for the sole reasons of time and hesitancy of giving out proprietary customer data.

Using this dataset will allow me to make a useful model for predicting customer retention, which is a big factor not just in the banking world, but also in retail, insurance, subscription-based products, etc. It can also be used as a reference to my employer if he ever decides that a predictive model would be advantageous in his business.

Problems that could arise might be that the model shows bias toward older people since the dataset is skewed in favor of people over 50. Another bias could arise with customers with a zero balance being weighted as more likely to churn since there are many more zero balances than any other value in that category. A possibility might be gender bias between males and females. Still, I cannot see that being more of a problem than the prior two examples since the distribution of the gender column is fairly evenly distributed.