There are two questions I want to answer. The first question is, “What are the best set of predictors excluding Sales from the different regions for Global Sales?” The second question is, “Can we have a higher degree of accuracy predicting Global sales through the results of the previous question or by using a machine learning method?”

My hypothetical client is any game developer. They would care about this particular issue because if their goal is to make a profit, they would know what kind of game they should make or who should publish their game. For an example, if Platform and Genre were found to be good predictors, a game developer would know exactly which Platform contributes the most to Global Sales and what kind of game generates the most sales. If their goal was simply to make money, it would be in their best interest to create a game with a specific genre in mind or on a specific platform. If Critic score was found to be a good predictor, a game developer could research the games whose critic score is high and get a general sense of how critics rate games. They would be able to use those features and implement them into their own games to maximize their sales. There would be similar reasoning for the remaining possible predictors.

I am going to use the publicly available dataset found at <https://www.kaggle.com/rush4ratio/video-game-sales-with-ratings>

My approach to this problem is simple. First clean the data, and then apply statistical methods to answer the first question. To answer the second question, I would divide the dataset into train and test, and then find how accurate the previous found model performs and then create a new model using machine learning. I would then compare how accurate this new model performs when compared to the previous model.

My deliverables are code, a report, and a jupyter notebook (with code, output, and a report).