Capstone Project 1 Proposal

I’m going to attempt to help solve the problem of selecting whether to undergo a real estate development or acquisition project. Using the rough information given to me by a self-storage company, I will analyze past project proposals and whether they are undertaken or not, to determine whether a future project should be deemed worthwhile.

By solving this problem, I will hopefully be able to significantly speed up the process of selecting which projects to take on. Many projects are proposed every week at an increasing rate, with the majority of them being turned down. Although this is the case, almost every project proposal has to be looked at extensively before it can be determined whether to turn them down or not. If by using this data I am able to determine more quickly which projects to take on then much time will be saved and more profit-generating projects can be accepted.

In the short-term, what I have listed is the goal for this problem, but in the long-term this solution can potentially offer even more. In the real estate self-storage business the ultimate profitability of a project cannot be determined for several years, so this analysis is only determining whether based on the history of the company whether a project should be considered or not. In several years, the profitability of projects will become available and using a very similar data set and model an estimate of the profitability of a project could be determined, offering a greater solution than what the current data can give.

To solve this problem, I initially will be given individual project proposal workbooks from the past four years. From this data I will use python to aggregate the relevant information in the project workbooks into a single workbook that will have the pertinent information for a given proposal in each row of the workbook. After compacting this information, I will clean and analyze the data. This consists of looking for extreme outliers, empty values, and making the data into a readily machine-readable format. Once the data has been fully cleaned, initial statistical analysis and EDA will be done. Reflecting on this analysis, the data may be cleaned again to determine if some columns of data may be removed or not, whether to fill missing values or not, and generally get the data into its desired shape. Once the data is considered fully prepared, I will begin using several machine learning models with many different parameters to determine the best model for the data.

In the end, I will deliver the code for the machine learning model, the new compact spreadsheet of historical project proposal data (with code to continuously update this sheet), and a written up summary of the project with analysis, suggestions, and discoveries along the way.