Can we use data collected from historical projects to model a project’s profitability with high precision? This company requested us to use this data in order to help them identify the metrics that are good predictors of project’s profitability. In addition, by identifying these features that are associated with a project’s profitability, they will be able to improve how they conduct future projects.

Criteria for a successful project: students produced results that are interpretable and can be used to make inferences on variables that are predictive of profitability. Students created a model that determines which variables impact profitability, and based on those variables, predict if a project in the future will be profitable or not profitable. Students determined the primary influencers of a project’s likelihood to succeed or fail. Students may identify what additional types of data might be needed to create reliable predictions. Students have identified the measures of the model’s precision/accuracy as well as noted the limitations of their analysis.

Walter P Moore receives contracts to do projects, which they price out using various methods. The employees work on these projects, or work may be subcontracted out for these projects, both which come at a cost.

The data are for 20k+ completed projects.

Profitability can be defined in many different ways, such as whether the project lost or did not lose money. Other profitability indicators can be whether the project gained a profit at some margin. The first 8 digits of the project code is the project family, and denotes a unique individual project. Each project may have sub-parts, which some sub-parts are profitable, while others are not; we are most interested in the overall project profitability as the measure for profitability. Note that not all sub parts of a project belong to the same project, but for the purposes of this challenge, we treat each unique first 8-digits of the project code as an individual project.

The data contain repeated measures for the variables, but the profitability is one measurement for the project. These repeated measures may not all be on the same time scale, so keep that in mind when manipulating the data and building your models.

Cost-plus projects are known to be profitable no matter what because they are billed by each hour of work at a profitable rate; it may be of interest to remove these projects from your data prior to analysis. Projects that start with project code “15” are not profitable because it is the billing code used for corporate overhead work; it may also be of interest to remove these projects from your data prior to analysis.

Walter P Moore has evaluated these data before using black-box methods with approximately 90% accuracy, but would like to have a more interpretable model that can be utilized in practice.