Problem Statement and Dataset Description

Team: Maddie Clubb, Danish Bokhari, Will Xue, Armando Mendez, Hussiny Mohammed

Problem Description

For our final project, we want to build a model that predicts the time between a job's completion and receiving the final payment for the job, for Imperial Granite and Marble. The problem is significant because family businesses need to be relatively conscious of how much money they have available to put back into the business – thus, knowing when payments are likely to come in for various jobs is important for managing budget and planning inventory purchases. There has been no previous attempt to solve this problem, past personal judgment.

The dataset we are going to use is from Armando's father's family-owned countertop-installation business: Imperial Marble and Granite (IMG). We got this data from Armando's family since it is their family business. We currently have 1400+ data points from 2019-2021, and if need be, we could obtain further data from previous years. The features we have available to us include:

- CONTRACTOR
- CUSTOMER NAME
- PO#
- STONE COLOR
- DATE INSTALLED
- PLACE INSTALLED
- SQ.FT
- PROJECT COST
- DEPOSIT AMOUNT
    - CHECK #
    - DATE
- FINAL PAYMENT
    - CHECK#
    - DATE
- PAYMENT DAYS OUTSTANDING
- TOTAL PAID

Additional features could be collected from dataset, including features from purchased inventory & remaining inventory data sheets. Model target will have to be constructed by subtracting Final Payment Date from Date Installed. Target seems to fluctuate between a few days and a little over a month.

In terms of the models we will explore, we want to implement either a regression model (perhaps SVM Regression) or a classification model using binned date intervals. After the training of our models, we plan on performing performance evaluations to find the best model using an 80:20 holdout ratio with 5-fold CV.