# **Project Status Report**

## Name: Zachary Tankersley

## Date: June 22nd, 2022

## Status Report

| **Status** | Progress |
| --- | --- |
| **Supporting Documentation** | * I wanted to eliminate unnecessary variables in order to provide a little bit more clarity of purpose and general simplicities sake. Although topographical information (family, soil information) isn’t strictly necessary, I decided to leave it in just to provide additional context or possible insight. |
| **Preliminary Data** | * The .CSV dataset file provided was downloaded and opened in Microsoft Excel, where I then removed the variable columns that were not needed, such as Elevation. Once those columns were removed, the “column edit” version was saved as a whole new spreadsheet document. The top row was then removed in order for the import process into the SQL database to work correctly. A check was also run at this stage in order to make sure each field was populated by the correct code and that no field was left blank. The table was then fed into the SQL Workbench tool via the import wizard, allowing me to verify all columns in the table matched all columns in the cleaned .CSV file. The contents of the table appear to be in usable order. |
| **Data Analysis Requirements** | * At face value, while performing some basic functions in the SQL Workbench tool, the data does appear that it should *mostly* work for our purposes. The data in the table are represented by code values which indicate different qualities or statuses of the trees. Some variables are represented by literal measurement values such as the variables indicating distance to water. Determining which areas are the highest priority will be somewhat more difficult depending on what the Forest Services deem to be more important. In the client account documentation, the areas of high-value are those of spruce/fir species that are also at high elevations, which this data does provide, however it is unclear if closer proximity to a water source or roadway increases the “value” of the area. The codebook provided isn’t necessarily formatted in a “sloppy” manner, however the lack of clarity, comma separation, and even outright errors in the original codebook (especially in regard to soil type information) leads me to believe that the data may not be as reliable as it may seem. |
| **Next Steps and Rationale** | * Additional considerations that the Forest Service had per their account information would be things like the effectiveness of insecticides, or pheromone packs for controlling the beetle populations, or most effective methods to thin sections of dense forest or removing underbrush and dead growth. These data may have to be gathered from many different sources such as agricultural studies or insecticide manufacturers. * Other than the “BLOB” field, the data does appear to be sufficient to answer the current basic research questions. * There are no restrictions on this data, it does not contain any data types that are protected or sensitive. |

## Supporting Documentation

In addition to the status report, be sure to submit a distilled version of the codebook to help your stakeholders clearly understand what preliminary data is available.

## Preliminary Data

Graphical user interface

Description automatically generatedGraphical user interface

Description automatically generated with medium confidence