# DAT 223 Project Two Client Proposal Part 2 Template

**Organization Problem for Western Forest Service  
Analyst’s Name: Zachary Tankersley**

**Date: June 21st, 2022**

## Research Questions

1. Which areas of the forest are the most important to focus efforts on?
2. What is the average distance from these critical areas to water sources?
3. Is there any significance in soil type similarities in these high-priority areas?

## Data Types

To determine which areas are the most important we will need to compare qualitative descriptive data. Soil type similarities are also descriptive and therefore qualitative. Distance to water is a measurement and therefore quantitative. Distance from roadways is also quantitative. Some of these data points may need to be cross-referenced to highlight the areas of highest priority if a scale or priority rating is desired.

## Success Criteria

1. Areas of critical interest have been identified.
2. Soil similarities in critical areas have been identified.
3. Areas near water *and* of high priority have been identified.
4. The average distances to roadways for critical areas have been identified.

## Variables of Interest or Units of Analysis

The Cover\_Type variable, will be of primary interest as it describes the tree types which can be used to eliminate irrelevant “trees” (Spruce trees are the focus). This variable will help pinpoint the areas considered to be a high priority when analyzed in conjunction with other variables such as Wilderness\_Area(1-4). The Horizontal and Vertical\_Distance\_To\_Hydrology variables will be useful in measuring the average distance from critical areas to a body of water. Horizontal\_Distance\_To\_Roadways variable will be useful for measuring the average distance to roadways. These are all independent variables as these traits are not influenced by other variables. The other variables included in the data set may have some possible influence on beetle population or site value, such as elevation, but those potential influences were not clearly defined by the forest service previously. For the qualitative data types, the units of analysis are down to individual trees as specified in the code book.

## Data Gathering and Analysis Approach

Once the .CSV file has been received the data will have to be cleaned and prepared. Any anomalous data entries need to be identified and corrected or adjusted as will any missing data. Once the data has been prepared it can then be imported onto a data analysis SQL server where it will then be imported into the SQL workbench tool to be sorted and analyzed. For any additional data or insight to be provided, such as the most effective control method for the beetle population, that data will have to be gathered, vetted, and fed through the same process to ensure accuracy.

## Rationale

The data provided provides us with several great variables or metrics that we can use to gain insight into the primary questions provided by the Forest Service. To answer the questions regarding distance to water and soil similarities, we must first identify and isolate the high-priority areas. As stated by the Forest Service, the Spruce Beetle population can infect *any* spruce tree within a geographical range, so we need to start the search by removing the irrelevant tree types via filtering the Cover\_Type variable and eliminating all non-spruce species from the data set. Afterward, we need only calculate the average distance from those trees left, and bodies of water reflected in the Horizontal or Vertical\_Distance\_To\_Hydrology as well as the average distance to roadways via the average of Horizontal\_Distance\_To\_Roadways. The data provided should be sufficient on a basic level to answer the research questions posed by the Forest Service. However, identifying the best beetle control methodologies, success rates, tree thinning strategies, etc. cannot be answered by this data and that would require separate data gathering methods in addition to this analysis.

## Approval and Authority to Proceed

Managers to complete this section, please leave blank.

We approve the project as described above, and authorize the team to proceed.

| Name | Title | Date |
| --- | --- | --- |
|  |  |  |