Group 2 - John Fowler, Nicholas Furlow, Anastasios Kydes, and Adrian Beldzinski

Professor Moore

OPIM 3211

7 May 2021

UCONN OPIM Group 2 Final Report

Empathize, Define, and Ideate

The organization that we were partnered with for the duration of this project was the *Town of Groton - Open Space Analytics,* an organization that is aiming to preserve open spaces in Groton. The stakeholder that we were in contact with throughout this process was Larry Dunn, a passionate individual who created a model that calculates the economic value of various undeveloped open spaces and then compares it to the cost of developing the land. In order to effectively convince the Town of Groton to implement his strategy of developing undeveloped open spaces in Groton, improvements and validation of his model and data collection methods needed to be made.

The current state of the model needed to be improved in order to work as solid evidence to convince the Town of Groton to implement this strategy. The model would need to be easily accessible and understandable to ensure that the committee for the Town of Groton would be able to understand it and realize the benefits of developing these underdeveloped areas. Another problem with the model was the lack of visualization within itself. Visualization tools allow for viewers of the model to easily see and compare differences within the model as well as allowing for an easier understanding of the data itself. The final problem that Mr. Dunn was facing and needed our assistance with, was the validation of his model that he created. The model would not

be able to be used as any evidence to support the benefits of developing these open spaces without being validated for accuracy. Without these main problems being solved of the model, it would not be able to effectively be presented to the Town of Groton committee and used to persuade them into implementing the current plan.

After meeting with Mr. Dunn, we were able to brainstorm a few opportunities that would help him improve his model and data. First, we would have to improve the usability and qualitative findings of the model. This would allow for easier data consumption for different audiences which will allow for more people to understand the basis of the model and the information it provides. Next, we would utilize graphs and macros within Excel to help create more visualization in the model. By using macros, we can allow users to enable and disable certain graphs to help limit the amount of confusion while also increasing visualization of the model. We were also able to use JMP to validate the accuracy of the model by showing comparisons as well as displaying the most deciding variables in determining the value of open spaces. By utilizing all of the technologies available to us, we were able to ensure that the model conveys the absolute benefit that developing undeveloped spaces brings to the individuals and towns alike.

Prototyping and Documentation

Through our discussions with Mr. Dunn, it was clear that enhancing the usability and visualization capabilities of the current model were important actions that needed to be taken. The original model in Excel was quite complex, with several sheets and many moving parts, such as cells that were combined using formulas which would then feed into more formulas which would then feed into outputs. The complexity resulted in a spreadsheet that was difficult

to use and tough derive conclusions from, which was its main purpose. We set out to make the spreadsheet more visually appealing and easier to use.

Before creating a new spreadsheet based on the one we were provided with, we brainstormed how to make things more efficient. In the model's original state, there were five sheets, one devoted to each of the model's main input variables. Within each of these five sheets, there were inputs, constants, and formulas specific to each variable. Additionally, each variable was then broken down by the type of land (open space, commercial, etc.). Thes net economic values for these individual land types were then aggregated by adding up each of their five input variables in a "Summary" sheet. We believed that a different way of organizing the information in the spreadsheet would be more efficient for inputting information and understanding how inputs and formulas flow through and affect the output of the model. Instead of having one sheet devoted to each variable, we thought it would be beneficial to create one sheet for each land type since, in the end, comparing land types through net economic value was the end goal of the model's outputs. Additionally, we noticed that there were similarities between certain variables across each land type, which boded well understanding variables relative to one land type and another. After this brainstorming breakthrough, we began to implement these changes in a new spreadsheet.

This new spreadsheet was developed in a new Excel document. The original spreadsheet was used as a template for correctly bringing over the formulas and constants in that model. Meticulous care was taken when transferring information between spreadsheets to ensure that every constant and formula was identical to the original and that nothing was lost or went awry. This was a time consuming process, but one that needed to be perfect in order to make sure that things worked as they should. The completion of this arduous step meant that all of the

information in the previous spreadsheet was retained and successfully reorganized in the new spreadsheet in order to make things more efficient. Now that each land type had its own sheet, it was clear to see how each variable for each land type was affected and how the overall net economic value for each land type was determined. In order to increase the usability of these sheets, two main things were done. First, the cells that listed the variables and output for each land type were locked at the top of each sheet so that they could be easily visible from any cell. Macros were also created to provide effective visualizations that clearly demonstrated the relationships between values in certain areas of the sheet. Both of these additions give the user more options and makes their experience more efficient.

In order to bring the information from these sheets into one area, a "Summary" sheet was created. This sheet contained two main areas. The first involved taking the net economic value from each land type, which was derived in each of their respective sheets, and listed them one next to the other. The second satisfied one of the main criteria that Mr. Dunn was looking for. He wanted a way to demonstrate the value of open spaces relative to other land types. Therefore, a bar graph was created that visually incorporated the information from the first area of this sheet. The open space bar was colored green while the other four were colored grey. A bar graph was the visualization of choice because of its effectiveness at displaying differences in numeric values as well as its ease of interpretability.

Additionally, to add a layer of protection, cells that were not classified as inputs were frozen through cell protection methods. This way, if someone maliciously attempted to change certain aspects of the spreadsheet that they should not be able to, they would not be able to. Password protection for the unlocking of this mechanism was asked about, but Mr. Dunn said that using the simple buttons on the Excel toolbar were enough for this feature. Data validation

rules were also enacted to ensure that only decimal or whole number values are allowed to be typed into input cells.

To round things off, a “Guide to Sheets” tab was created. In this sheet, each subsequent sheet was listed, including a hyperlink to that respective sheet, along with a description of the land type, or information, that it was responsible for. This sheet was placed at the beginning of the Excel file and provided a nice introduction to the spreadsheet along with tools for efficient navigation between sheets. Additionally, instructions for protecting and unprotecting the spreadsheet were included in this Guide. Finally, one cell was highlighted in green and labelled “inputs” and another was labelled grey and labelled “constants” so that certain cells and their functions could be easily identified.

From our initial conversations with Mr. Dunn, we came away understanding the importance of data validation in our task. Thus we set out to not only transform his model into something more easily used, but also determine if it was truly useful in the first place, and to potentially uncover any upgrades that could be made to increase veracity of the model.

To do this, we first poured over the source documentation Mr. Dunn had provided for us. We found it important to decide if his methods were logical, as an evaluation of property in this manner can be nebulous and incredibly subjective. Logically speaking, however, most of his conclusions and calculations used by the model made sense. Furthermore, the sources he utilized for making these determinations were sound. We attempted to find any newer, more contemporary studies that could potentially update existing figures in the model, yet were unable to uncover any.

Moving forward from here, we constructed another Excel spreadsheet, this one with the purpose of being transformed into a dataset for eventual statistical analysis. This model would be used to evaluate the Open Space Economic Value for real open spaces in Groton. We adapted the various calculations and inputs from Mr. Dunn's existing model into columns in this new datasheet, as well as added other inputs for things like property value and gross assessed value, should anyone wish to compare these values to the OSEV.

In order to utilize real-world open space data from Groton, and to maintain a sort of consistency, we decided to utilize sources that Mr. Dunn himself utilizes, with the most notable source being the town of Groton ArcGIS Viewer, where property information could be obtained with ease and accuracy.

At the onset of this model's creation, we already discovered an area that could utilize improvement: the pollutant cost variable's calculation. In Mr. Dunn's original model, we found it was lacking in specificity, only calculating the positive impact trees can have on pollutants under several circumstances. We, as well as Mr. Dunn, thought it would be appropriate to improve upon this by incorporating a new way of evaluating this cost.

We developed a sort of ranking system for this purpose. Open spaces that did not fall into those with predetermined pollutant cost value would be given a rank from 0-3, with each increasing level denoting an increase in average tree density per acre. A rank of "0" indicates less than 10 trees per acre, "1" indicates up to 40, "2" indicates up to 80, and "3" indicates up to 120. (Figure 1.4) These amounts may seem arbitrary, but as far as we can tell are in line with the average density of forested areas in Connecticut. This concept was incorporated into the model

in the form of the Tree Density input. The density of a specific open space can be determined by use of a tool like Google Maps, utilizing a satellite image.

We generated 30 rows of data from various parks and open spaces around Groton, as this would grant statistical normalcy, and give a wide enough range of types of open spaces to analyze. This data collection took an extremely long time to complete. Certain types of data had to be generalized, such as average property values surrounding an open space, as we lacked the resources to analyze this information efficiently and effectively.

After fully integrating the data from the GIS website, we decided to analyze this information for any notable patterns utilizing an external piece of software. In this case, we utilized JMP for this analysis, as integrating Excel files would be easy.

Within the JMP file, we first checked the OSEV column for any outliers, in order to determine if there were any extreme differences in how this value is calculated. We found several outliers, with one outlier being particularly egregious, the one for the Copp Property. (Figure 1.5)

Next, we ran a basic linear regression to assess the impact of the variables within the equation, and see if any were more impactful than others. Our analysis showed that the variables were fairly evenly impactful, though some do outweigh others in its current state. For instance, the LogWorths of Natural Systems Value, Pollutant Cost, and Tax Revenue are all about the same, while the LogWorth of Public Service Cost is noticeably lower. For this reason, we believe that Mr. Dunn should reweigh the values in the model accordingly, potentially by decreasing the emphasis on Public Service Cost. This regression analysis also contained other smaller variables to see if any were being overlooked, but their low LogWorth values indicate this is not the case.

This model also contains a sensitivity analysis at the bottom, where an individual can analyze the changes in other variables by the alteration of one variable. (Figure 1.6)

Our JMP file also contains a Multivariate model of every variable within the dataset if Mr. Dunn desires insight to specific variable correlations, if he wishes to reweigh finer aspects of the overall model. Furthermore, we also generated a decision tree model and a corresponding leaf report to do more variable analysis from another perspective. In this, we found the model tends to split most frequently on Natural Systems Value, (Figure 1.7) emphasizing that variable over others. This information can also be utilized in recalibrating variable weights.

Ultimately, if there is one aspect that should be recalibrated, we feel that it would be most effective for Mr. Dunn to place a decreasing emphasis on acreage of land as that amount increases. We feel that the most notable outliers were also the largest parks, as their natural systems value was being inflated by this massive acreage. We feel that incorporating weights to reduce the impact would improve the accuracy of the model incredibly.

In terms of documentation, we included two descriptive sheets on the Excel dataset informing a user how to interpret certain columns, the purpose of the document, and where and how to input values in certain columns. Columns are also color coded for user convenience, with green columns denoting user inputs, yellow corresponding to one of the main component outputs, orange corresponding to the final open space economic value, and white are simply calculations. The model is ready for further use at Mr. Dunn's discretion for further analysis. Documentation for accessing our models within the JMP file is also provided here.

In conclusion regarding this model, we feel satisfied in our ability to guarantee the veracity of Mr. Dunn's model as it is, and we feel that our recommendations would only make it more effective.

Feedback

Throughout our time working on the Open Space model, we have had a focus on continuously building on the foundation of the initial file. Upon receiving feedback from the prototype report, we realized the details our team needed to focus on pertained to:

- **Uniformity** primarily within the input tabs that the user would be frequently utilizing
- **Summary tab** that outlines the Net Economic Value as it is variable normally used for comparing costs of the various areas
- **Worksheet protection** to disallow users from accidentally changing the constraints of the model

To enhance the uniformity of the model, we concentrated on elements that we knew the user would use most. Specifically, we paid close attention to the new Macro buttons used. Outlined in the graphic below are the various button placements used within each tab that is primarily for user input. The size of the buttons denotes its purpose, where the smaller buttons convey statistics of one variable, namely the column that they are above, while the larger ones cover the entirety of the table associated with the input.

The following are the purposes of the highlighted buttons within a table. An example for each number is denoted within Appendix Figure 1.1:

1. **Variable Buttons** – When clicked upon, a bar chart for the selected variable will be created reflecting the new input
2. **Clear Buttons** – When clicked upon, the associate chart is removed
3. **Summary Button** – When clicked upon, a summary of the information within the adjacent table is created; the 'clear' button removes the created chart (an example is shown in Appendix Figure 1.2)
4. **NSV Summary Button** - When clicked upon, a summary of the information within the adjacent table is created; the 'clear' button removes the created chart (an example is shown in Appendix Figure 1.2)

Another element added to the model post prototype meeting was the addition of the 'Summary' tab, denoted in Appendix Figure 1.3. Initially, in order to compare the Net Economic Value, the user would have to cycle through each tab and compare the different amounts individually. The 'Summary' tab now allows for a central place within the model to observe the variable across all the tabs at once.

The last added element post prototype meeting was the enablement of worksheet protection across all of the tabs. This addition intends to limit the possibility of a user changing a constraint by accident, which, if done unintentionally, could break the model entirely. Since the protections are not associated with a password, the user may take the protections off when needed.

Reflection

Throughout our work on the Open Space model, we have used a variety of tools within Microsoft Excel and JMP.

In the case of Microsoft Excel, we focused on tools that would help improve the model's usability. Specifically, we used Microsoft Excel's macros capability to record a series of actions that translated to the creation of visual charts based on the user's input. The intention behind recording these macros is to enhance the user's ability to create visualizations based on their input by clicking the desired buttons on the developmental area tabs. Additionally, we created a tab named "Guide to Sheets" with a list of hyperlinks within the model to allow for the user to quickly access the desired tab without scrolling through the bottom of the document.

In the case of JMP, we focused on analyzing the model's validity. The tools used within the application include the Variable Analysis, Leaf Report, and Decision Tree. The focus of using each tool was to identify any outliers within the model that could potentially be decreasing its validity.

During the beginning of our project, we went about creating a Gantt chart (Figure 1.8) to help keep us organized and working at a steady pace during the duration of the semester. However, as time progressed, we realized that we were unable to keep up with the chart set up for us. This led to us slowly falling behind our originally intended plan until we figured out that we would be able to do the validation of the data at the same time as we were reorganizing the mode instead of after. We divided up the work and were able to catch back up and complete both the model redesign as well as the validation of the data. This was a good learning experience for us because we were able to make adjustments as a team when we realized that we were unable to keep up with the original plan set up for us. Sometimes things don't go exactly as planned and therefore being able to adjust and make decisions on the fly as a team is important to have.

With the application of the different tools used, we can confidently say that our knowledge within the applications has grown vastly. Prior to the project, we had worked within both Microsoft Excel and JMP for smaller short-term assignments, however, this semester-long project challenged us to put our individual knowledge/experience together to continuously build onto the model week after week. This timeline allowed us to go past our comfort zones within each application and utilize tools that we were not as familiar with prior to, such as macros within Microsoft Excel and in-depth validation analysis within JMP.

Appendix

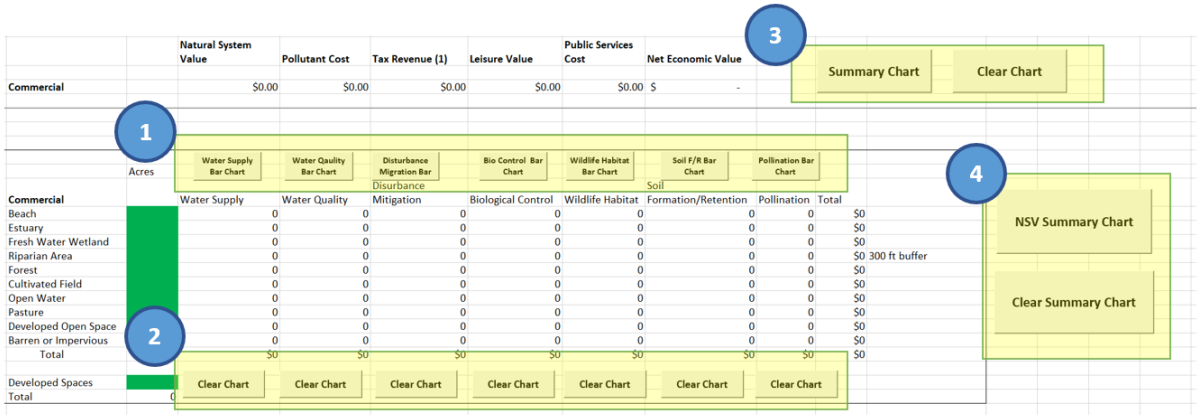


Figure 1.1: Commercial Table Buttons

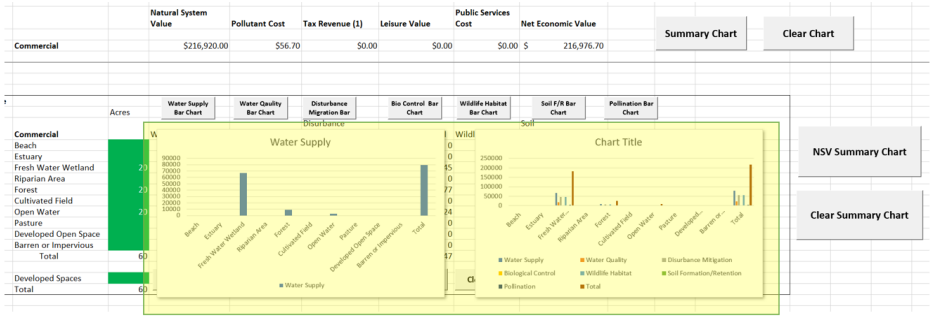


Figure 1.2: Commercial Table Bar Charts

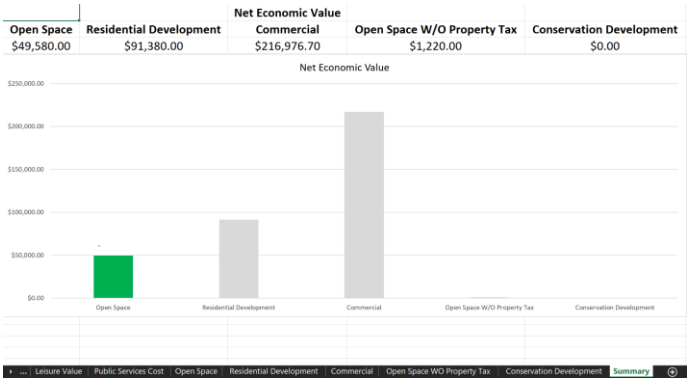


Figure 1.3: Summary Tab Bar Chart

Tree Density
0 - Little/No Coverage (up to 10 trees/acre)
1 - Low/Medium (up to 40 trees/acre)
2 - High (up to 80 trees/acre)
3 - Extreme/Forest (up to 120 trees/acre)

Figure 1.4: Tree Density Breakdown

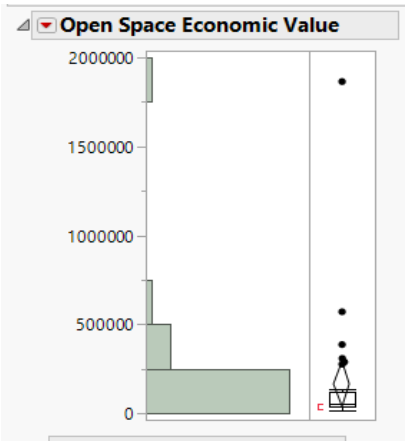


Figure 1.5: Outlier Analysis

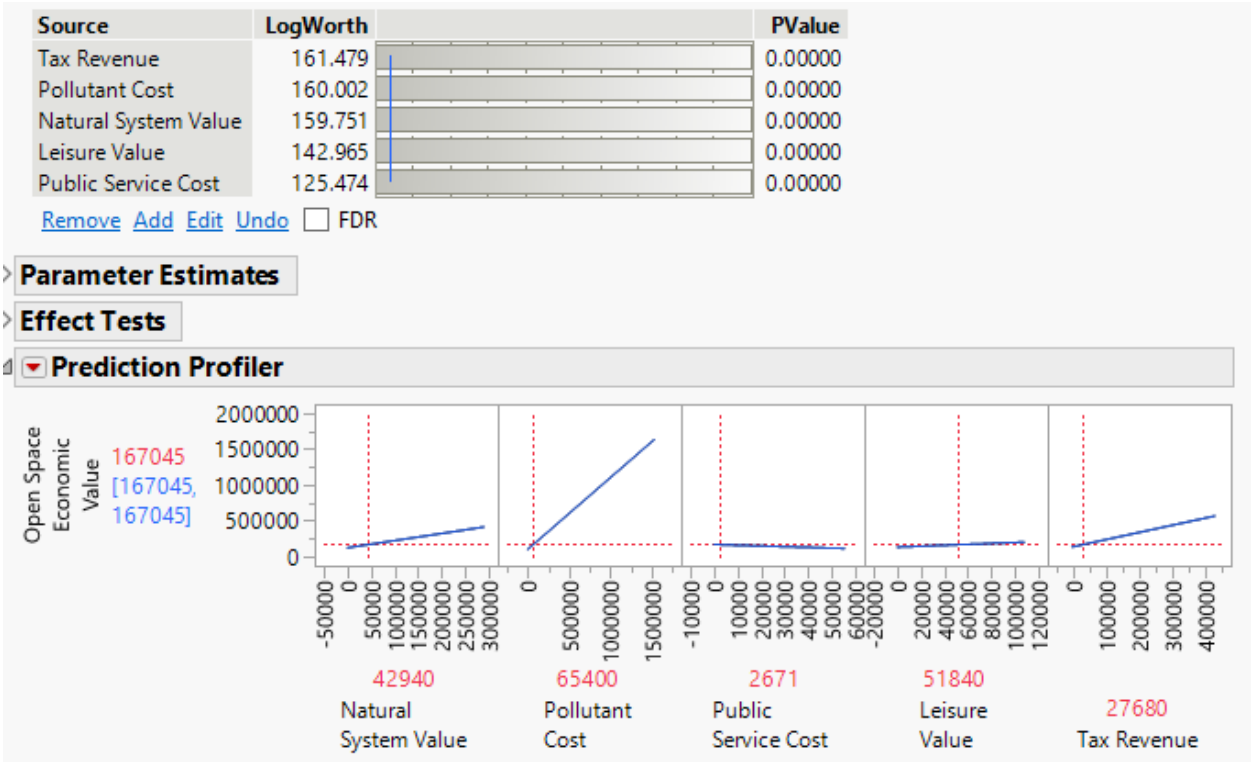
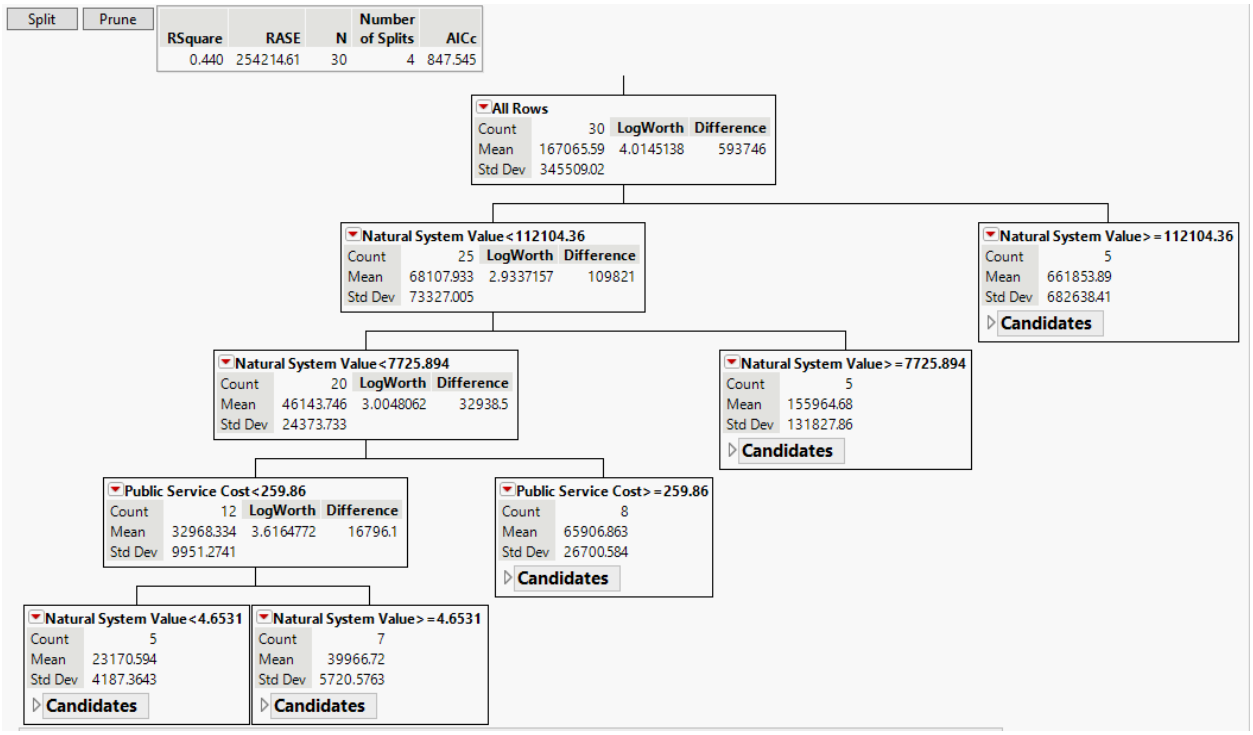


Figure 1.6: Regression Analysis



Leaf Report

Leaf Label	Mean	Count
Natural System Value<112104.36&Natural System Value<7725.894&Public Service Cost<259.86&Natural System Value<4.6531	23170.594	5
Natural System Value<112104.36&Natural System Value<7725.894&Public Service Cost<259.86&Natural System Value>=4.6531	39966.72	7
Natural System Value<112104.36&Natural System Value<7725.894&Public Service Cost>=259.86	65906.8625	8
Natural System Value<112104.36&Natural System Value>=7725.894	155964.684	5
Natural System Value>=112104.36	661853.886	5

Figure 1.7: Decision Tree and Leaf Report

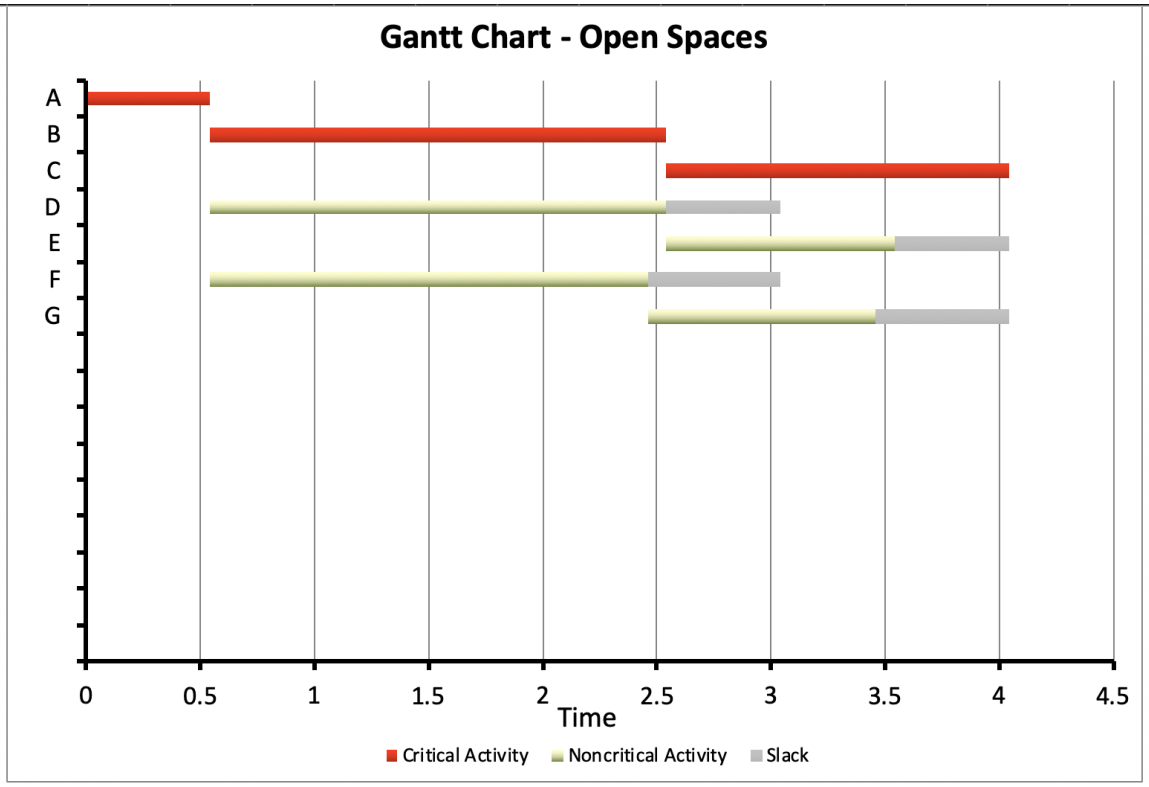


Figure 1.8: Gantt Chart