Data Analysis of Environmental Parameters of Soil

Data-151: Intro to Data Science

By: Jack Colwell, Owen Galicia, Santiago Gutiérrez-Morales, and Lilu Smith

I have neither given, or received, nor tolerated others’ use of unauthorized aid.

* *Jack Colwell, Owen Galicia, Santiago Gutiérrez-Morales, and Lilu Smith*

1. Proposal

This project will analyze data from the previous research projects: *Environmental Restoration Target Estimation Around Engquist Nature Preserve* and *Assessing Carbon Sequestration Potential of Cropland Conversion in Porter County, IN*. These research projects were done at Shirley Heinze Land Trust (SHLT) by students of the Geography Department at Valparaiso University, including Jack Colwell (member of this group), Korbin Opfer (current student), Doc Janowiak (recent alumnus), and Justin Self (recent alumnus). This research is also supported by Dr. Jon-Paul McCool of the Geography Department. Additionally, this research is still ongoing this semester. These past projects collected soil samples from different environment types at properties owned by SHLT and the samples were subsequently analyzed for various environmental parameters. Environmental parameters that were specifically analyzed were bulk density, gravel mass, rock volume, soil volume, wet weight, dry weight, mineral sample, organic matter percent, soil organic carbon percent, calcium carbonate percent, estimated carbon in horizon per sq. m., and estimated carbon in 10 cm per sq m. This project will explore healthy ranges for environmental variables for agriculture and then compare that to our existing dataset. We hope to answer the question: which land cover type is the best for agriculture based on certain environmental parameters? Through this we plan to run different analyzation techniques, like a correlation matrix of the different parameters created taught in our Data-151 class and other predictive models. This would then show us if one soil parameter can accurately predict another, or at least the magnitude and direction of another parameter. Baseline soil metrics are currently being researched by our Data-151 Team to test our data against. The answer to this question can help farmers identify prime real estate for new agriculture development.

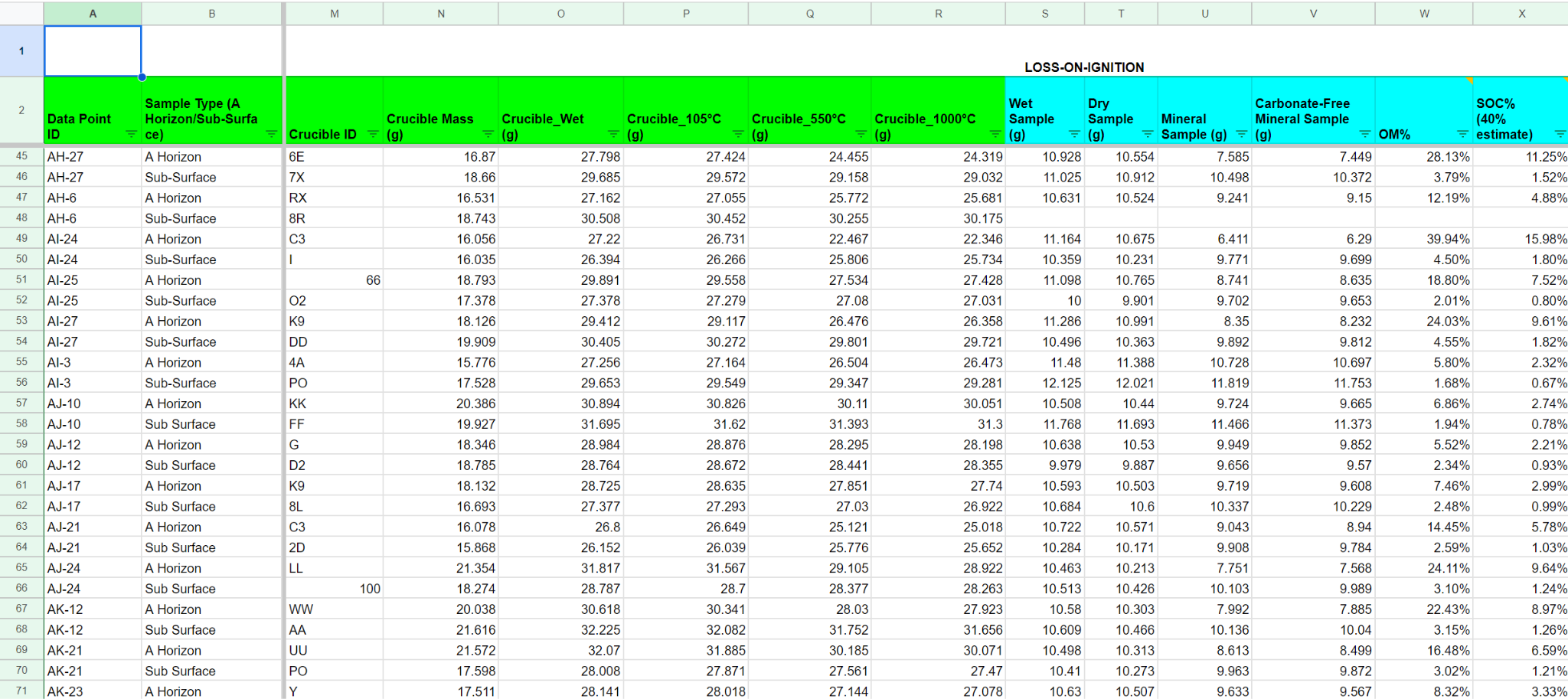


Fig. 1: A snippet of the data that was collected and analyzed from the aforementioned research projects.

1. Assessment

Our research question is what land cover type is best for agriculture based on certain environmental parameters. Our H0 (null) is: There is no difference between what land cover type is the best for agriculture. Ha: (alternative): At least one of the land cover types is significantly better or worse for agriculture than the other land cover types. Our hypothesis is important because it allows the researchers (us) to statistically analyze our data and draw accurate conclusions about whether our results are likely due to chance or an actual phenomenon. We will be able to tell if our research question is correct based upon researching what parameter values are adequate for agriculture, and seeing how many land types are able to achieve those numbers. We will be able to tell if our hypothesis is correct if our results are statistically significant. If they are, then our hypothesis (alternative) is correct.

1. Schedule

| Date | Description |
| --- | --- |
| 9/16/2024 | All members of the team will work on the Project proposal and presentation. All will oversee and edit the documentation   * Jack Colwell will write the Project Description * Lilu Smith will create a schedule * Owen Galicia and Santiago Gutierrez will write up the assessment and create the presentation |
| 9/20/2024 | Continue working on project proposal and presentation in class |
| 9/22/2024 | Clean up project and presentation to turn in |
| 9/23/2024 | EVERYTHING DUE  Presentations conducted during class |
| Week of 9/23/2024 | Start Part 2   * Use Jack’s data * Clean up the data (get rid of outliers) * Exploratory analysis of data * Deep data analysis of data |
| 9/30/2024 | EXAM DAY |
| Week of 9/30/2024 | * Figure out what predictive modeling is best for our project * Documentation of what has proved harder than expected, revisions you made to the modeling, what you have learned about your data and the topic in general |
| Week of 10/7/2024 | * Work on report * Work on short presentation with highlights of your exploratory data analysis |
| Week of 10/14/2024 | Clean up project and presentation to turn in |
| 10/18/2024 | EVERYTHING DUE  Presentations conducted during class |
| Week of 10/21/2024 | Start Part 3   * Draft the model * Research paper |
| Week of 10/28/2024 | * The draft of the final report should primarily focus on the introduction, data and methods, and results sections * Work on demonstration of your draft modeling efforts |
| Week of 11/4/2024 | Clean up paper and presentation to turn in |
| 11/13/2024 | EVERYTHING DUE  Modeling demonstration conducted during class.  Peer review of draft paper conducted during class on November 20, 2024. |
| 11/18/2024 | Start Part 4   * Complete research paper * Presentations |
| 11/22/2024 | EXAM DAY |
| Week of 12/2/2024 | WORK DAYS   * Research paper * Presentations * Clean up documents |
| 12/12/2024 | FINAL PAPER DUE  Presentations conducted during the final exam time for the course. |