**Lab Report**

Title: Lab 1

Notice: Dr. Bryan Runck

Author: Jarred Paquin

Date: 10/6/2021

**Project Repository:**https://github.com/jarredpaquin/GIS5571

**Google Drive Link:** N/A

**Time Spent:** 6 hours

**Abstract**

I set out to create an ETL pipeline that would take data from the MN Geospatial commons and combine it together using the Arcpy module. I was able to create the ETL pipeline and successfully downloaded two shapefiles. However, my spatial join failed and I was not able to produce the desired results.

**Problem Statement**

The goal of this lab was to create an ETL pipeline that could take data from the MN Geospatial commons and use the Arcpy module to combine that data together into one product.

**Input Data**

For this lab I used two shapefiles both created and maintained by Metro Transit. One was the transit routes in the region and the other was the transit stops. I chose these two datasets because they were thematically related. See Table 1 for more details.

Table 1. Data Used

|  |  |  |  |
| --- | --- | --- | --- |
| **#** | **Title** | **Purpose in Analysis** | **Link to Source** |
| 1 | Transit Routes | Test file to demonstrate ETL pipeline. | [MN Geospatial Commons](https://gisdata.mn.gov/dataset/us-mn-state-metc-trans-transit-routes) |
| 2 | Transit Stops | Test file to demonstrate ETL pipeline. | [MN Geospatial Commons](https://gisdata.mn.gov/dataset/us-mn-state-metc-trans-transit-stops) |

**Methods**

My methods closely mirrored those demonstrated in lab. I started by assigning my search query to a variable. I then navigated through the various lists and dictionaries to access the URL for the file I wanted to download. I then used the zip module to download that data as a zip file, and then unzipped it into my workspace. From there, I used the Arcpy module to change the spatial reference and attempt a spatial join, though unfortunately my join did not finish running and did not produce results.

**Results**

I was not able to get the desired results, as my join was not running correctly.

**Results Verification**

Unfortunately, I was not able to get my join to work, so I know that my final product did not succeed. If I had been able to get it to work, I would have verified that by checking the attribute table of the final shapefile. If it has attributes from both source datasets, then I would have known it succeeded.

**Discussion and Conclusion**

I found it a bit tedious to use an ETL pipeline to download data from the MN Geospatial Commons. It took significantly longer than downloading the data normally, and due to how unique the path was, it is not easily automatable to any other datasets. I did, however, find use in familiarizing myself more with the Arcpy module, as manipulating my data using Python has a lot of potential and use.

**References**

None

**Self-score**

|  |  |  |  |
| --- | --- | --- | --- |
| **Category** | **Description** | **Points Possible** | **Score** |
| **Structural Elements** | All elements of a lab report are included **(2 points each)**:  Title, Notice: Dr. Bryan Runck, Author, Project Repository, Date, Abstract, Problem Statement, Input Data w/ tables, Methods w/ Data, Flow Diagrams, Results, Results Verification, Discussion and Conclusion, References in common format, Self-score | 28 | 28 |
| **Clarity of Content** | Each element above is executed at a professional level so that someone can understand the goal, data, methods, results, and their validity and implications in a 5 minute reading at a cursory-level, and in a 30 minute meeting at a deep level **(12 points)**. There is a clear connection from data to results to discussion and conclusion **(12 points)**. | 24 | 16 |
| **Reproducibility** | Results are completely reproducible by someone with basic GIS training. There is no ambiguity in data flow or rationale for data operations. Every step is documented and justified. | 28 | 21 |
| **Verification** | Results are correct in that they have been verified in comparison to some standard. The standard is clearly stated **(10 points)**, the method of comparison is clearly stated **(5 points)**, and the result of verification is clearly stated **(5 points)**. | 20 | 5 |
|  |  | 100 | 70 |