**Lab Report**

Title: *Lab 1*

Notice: Dr. Bryan Runck (cc: Michael Felzan)

Author: Mattie Gisselbeck

Date: September 21, 2022

**Project Repository:** [Lab 1](https://github.com/mgisselbeck/GIS5571/tree/main/Lab1)

**Google Drive Link:** N/A

**Time Spent:** 20 hours

**Abstract**

The main objective of this project is to compare the different web APIs (Minnesota Geospatial Commons, Google Places, and NDAWN) using a ETL pipeline. The data used for the pipeline analysis will be specific to each of the online interfaces. All the web APIs will be analyzed in a python notebook in ArcGIS Pro using both open source and Arcpy tools. The results are qualitatively backed by the in-depth explanation of the analysis in the data flow diagram.

The results were uniformly showed by a printed data frame within a Python notebook or as points on a map. While the results were similar, the process of building a customized pipeline for each of the APIs can go many ways.

**Problem Statement**

The main objective of this project is to compare the different web APIs (Minnesota Geospatial Commons, Google Places, and NDAWN) using a pipeline that downloads two data sets, transforms both datasets to the same coordinate reference system (geographic and projected), spatially joins them, prints to the screen head showing the merged attributes, and saves the integrated dataset into a geodatabase (Runck, 2022). The interfaces of Minnesota Geospatial Commons, Google Places, and NDAWN use different spatial web APIs which requires us to build unique pipelines to extract its data.

*Figure 1. Building Pipelines for Google Places, Minnesota Geospatial Commons, and NDAWN*

(Note: This figure illustrates a simplified version of the constructed pipeline)

*Table 1. Required Data for ETL Pipeline Analysis*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **#** | **Requirement** | **Defined As** | **(Spatial) Data** | **Attribute Data** | **Dataset** | **Preparation** |
| 1 | Dataset from Minnesota Geospatial Commons | Raw Input for ETL Pipeline Analysis | N/A | N/A | N/A | N/A |
| 2 | Dataset from Minnesota Geospatial Commons | Raw Input for ETL Pipeline Analysis | N/A | N/A | N/A | N/A |
| 3 | Dataset from NDAWN | Raw Input for ETL Pipeline Analysis | N/A | N/A | N/A | N/A |
| 4 | Dataset from NDAWN | Raw Input for ETL Pipeline Analysis | N/A | N/A | N/A | N/A |
| 5 | Dataset from Google Places | Raw Input for ETL Pipeline Analysis | N/A | N/A | N/A | N/A |
| 6 | Dataset from Google Places | Raw Input for ETL Pipeline Analysis | N/A | N/A | N/A | N/A |

**Input Data**

The table below is a collection of data from Minnesota Geospatial Commons, North Dakota Agricultural Weather Network (NDAWN), and Google Places. The data will be used in the analysis and construction of a ETL pipeline for downloading into ArcGIS Pro via a Python notebook. Two datasets from each website were collected for spatial join, combining coordinate reference system, printing the joined table, and creating a geodatabase for the joined datasets.

*Table 2. Input Data (Minnesota Geospatial Commons, NDAWN, and Google Places)*

|  |  |  |  |
| --- | --- | --- | --- |
| **#** | **Title** | **Purpose in Analysis** | **Link to Source** |
| 1 | Assessed Waters, Minnesota, 2016 | Raw Input for ETL Pipeline Analysis from Minnesota Pollution Control Agency (MPCA) | [Minnesota Geospatial Commons](https://gisdata.mn.gov/dataset/env-assessed-water-2016) |
| 2 | Assessed Waters, Minnesota, 2022 | Raw Input for ETL Pipeline Analysis from Minnesota Pollution Control Agency (MPCA) | [Minnesota Geospatial Commons](https://gisdata.mn.gov/dataset/env-assessed-water-2022) |
| 3 | Monthly Weather Data, Baker, Minnesota, 2022 | Raw Input for ETL Pipeline Analysis from NDAWN | [NDAWN Center](https://ndawn.ndsu.nodak.edu/get-table.html?station=9&variable=mdmxt&variable=mdmnt&variable=mdavt&variable=mdbst&variable=mdtst&variable=mdws&variable=mdmxws&variable=mdsr&variable=mdapet&variable=mdtpet&variable=mdr&variable=mddp&variable=mdwc&year=2022&ttype=monthly&quick_pick=1_m&begin_date=2021-10&count=12) |
| 4 | Monthly Weather Data, Baker, Minnesota, 2019 | Raw Input for ETL Pipeline Analysis from NDAWN | [NDAWN Center](https://ndawn.ndsu.nodak.edu/get-table.html?station=9&variable=mdmxt&variable=mdmnt&variable=mdavt&variable=mdbst&variable=mdtst&variable=mdws&variable=mdmxws&variable=mdsr&variable=mdapet&variable=mdtpet&variable=mdr&variable=mddp&variable=mdwc&year=2022&ttype=monthly&quick_pick=&begin_date=2019-10&count=12) |
| 5 | Nearby Search, Google Places | Raw Input for ETL Pipeline Analysis from Google Places | [Google Places](https://www.google.com/maps/search/resturants+near+by/@44.9067304,-93.4077598,13z/data=!3m1!4b1) |
| 6 | Nearby Search, Google Places | Raw Input for ETL Pipeline Analysis from Google Places | [Google Places](https://www.google.com/maps/search/resturants+near+by/@44.9067304,-93.4077598,13z/data=!3m1!4b1) |

**Methods**

*Figure 2. Google Places*

*Graphical user interface, text, application

Description automatically generated*

*Graphical user interface, application

Description automatically generated*

*Figure 3. Minnesota Geospatial Commons*

*Application

Description automatically generated with medium confidence*

*Diagram

Description automatically generated*

*Figure 4. NDAWN*

*Text

Description automatically generated*

**Graphical user interface, application

Description automatically generated**

**Results**

*Map 1. Google Places*

*Map

Description automatically generated*

*This map shows the combined datasets of Nearby Search from Google Maps. All the points on the map represent a restaurant nearby the chosen coordinates. Also, this map helps to visualize the data results in the data frame table.*

*Map 2. Minnesota Geospatial Commons*

*Map

Description automatically generated*

*This map shows the combined datasets of assessed lakes in Minnesota from 2016 and 2022 from Minnesota Geospatial Commons. The blue points on the map represent lakes that have been assessed in Minnesota.*

*Map 3. NDAWN*

*Map

Description automatically generated*

*This map shows the combined datasets of monthly weather in Becker, North Dakota from 2019 and 2021. The weather symbol represents the coordinate location of the reported weather.*

**Results Verification**

The results are qualitatively backed by the in-depth explanation of the analysis in the data flow diagram. Also, the results were routinely checked to ensure correct outcomes and avoid an out of sequence data structure.

**Discussion and Conclusion**

Despite my beginner coding status, having exposure to intermediate-level coding helped me to dive into the deep end and learn a lot. The plunge was the catalyst to achieving all the project’s objectives and deliverables.

This lab has helped me build my confidence in learning how to code even if it means messing up 100 times before succeeding. The main take-a-way of this lab is how important it is to learn your style of coding and succeed by failing.

**References**

Runck, Bryan. *GIS 5571: Lab 1*. 2022.

<https://docs.google.com/document/d/1hREy9BSfYNHWedRdvJpk8U8ZD3zGs8cqlgOKioYMsZM/>

**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 | **24** |
| **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 | **28** |
| **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 | **20** |
|  |  | 100 | **100** |