

Lab Report

Title: Finding Coffee Shops That Are in the MPCA Sites Around My Home Using APIs (Lab 1)

Notice: Dr. Bryan Runc

Author: Mohsen Ahmadkhani

Date: 9/29/2021

Project Repository: <https://github.com/mohsen-gis/GIS5571.git>

Google Drive Link: -

Time Spent: 15 hours

Abstract

A Minnesota Pollution Control Agency (MPCA) site is a station that monitors and controls the environmental quality elements in the state of Minnesota. In this dummy project, to complete my PhD dissertation/program, I have applied and been offered a position as a geography research assistant to work in an MPCA station in the state having full access to their database. To start the program, I had to choose where to work. To do so, I had three criteria to choose based on. First, I want to work in a coffee shop. Second, I needed the coffee shop to be close to my home. Third, I preferred it to be near an MPCA site. In this project, I used google places API, and MPCA data from Minnesota Geospatial Commons API to download coffee shop and MPCA sites data. I used these two datasets to run the spatial analysis and all the required analyses. Also, to accomplish the other goals of the Lab1 assignment I downloaded and visualized the meteorological data from NDAWN dataset.

Problem Statement

In this dummy project, to complete my PhD dissertation/program, I have applied and been offered a position as a geography research assistant to work in an MPCA station in the state having full access to their database. To start the program, I had to choose where to work. To do so, I had three criteria to choose based on. First, I want to work in a coffee shop. Second, I needed the coffee shop to be close to my home. Third, I preferred it to be near an MPCA site.

Table 1. The list of required data sets for the proposed study.

| # | Requirement | Defined As | (Spatial) Data | Attribute Data | Dataset | Preparation |
|---|--------------------------------|--|----------------------|----------------------------|---------------|---|
| 1 | Coffee shops in Falcon Heights | All the coffee shops within the 5km neighborhood of my place | Point data (Lat/Lng) | Name, rating, address, etc | Google Places | Download using REST API as a pandas df and convert it to GeoPandas df |
| 2 | MPCA Sites in St. Paul | A site is a location, such as a manufacturing plant or cleanup site, where the MPCA carries out an activity. | Point data (Lat/Lng) | ActiveSite, cityName, etc | mnGeo Commons | Download using API, convert to GeoDataFrame and then |

| | | | | | | |
|---|---------------------|--|----------------------|----------------------------|-------|---|
| | | | | | | convert to polygon data |
| 3 | Meteorological data | A set of sites collecting weather data | Point data (Lat/Lng) | Temperature, humidity, etc | NDAWN | Download using REST API as a pandas df and convert it to GeoPandas df |

Input Data

To address the stated problem in this lab assignment, I used two different sets of input data. First, the coffee shop point data. I used google places API to download the data. This dataset contains all coffee shop locations as well as other attributes like name, rating, address, etc. The location of each coffee shop is stored as latitude and longitude. Second, MPCA sites data from Minnesota Geospatial Commons using API. This dataset contains attributes like ActiveSite, cityName, etc as well as latitude and longitude of each site. Also, I downloaded meteorological data from NDAWN API to satisfy the requirements of the lab assignment. Table 2 shows a list of these data.

Table 2. Input data description.

| # | Title | Purpose in Analysis | Link to Source |
|---|--------------------------------|--|----------------|
| 1 | Coffee shops in Falcon Heights | To find cafes that are in MPCA vicinity | API |
| 2 | MPCA Sites in St. Paul | To make it vicinity polygons to run spatial join | API |
| 3 | Meteorological data | For demonstration of NDAWN API usage | API |

Methods

The analyses in this project are divided in three groups. First, for downloading data I used the `urllib.parse` module to set the parameters for creating the API calls. I also used `requests` module to fetch data from the API provider sites. Second, processing the data, setting projection systems, building buffer, and running spatial join analysis I applied the `shapely.geometry` and `geopandas` packages. Lastly, to visualize the data and results I used `Folium` and `matplotlib` packages. The data flow diagram of this project is depicted in figure 1.

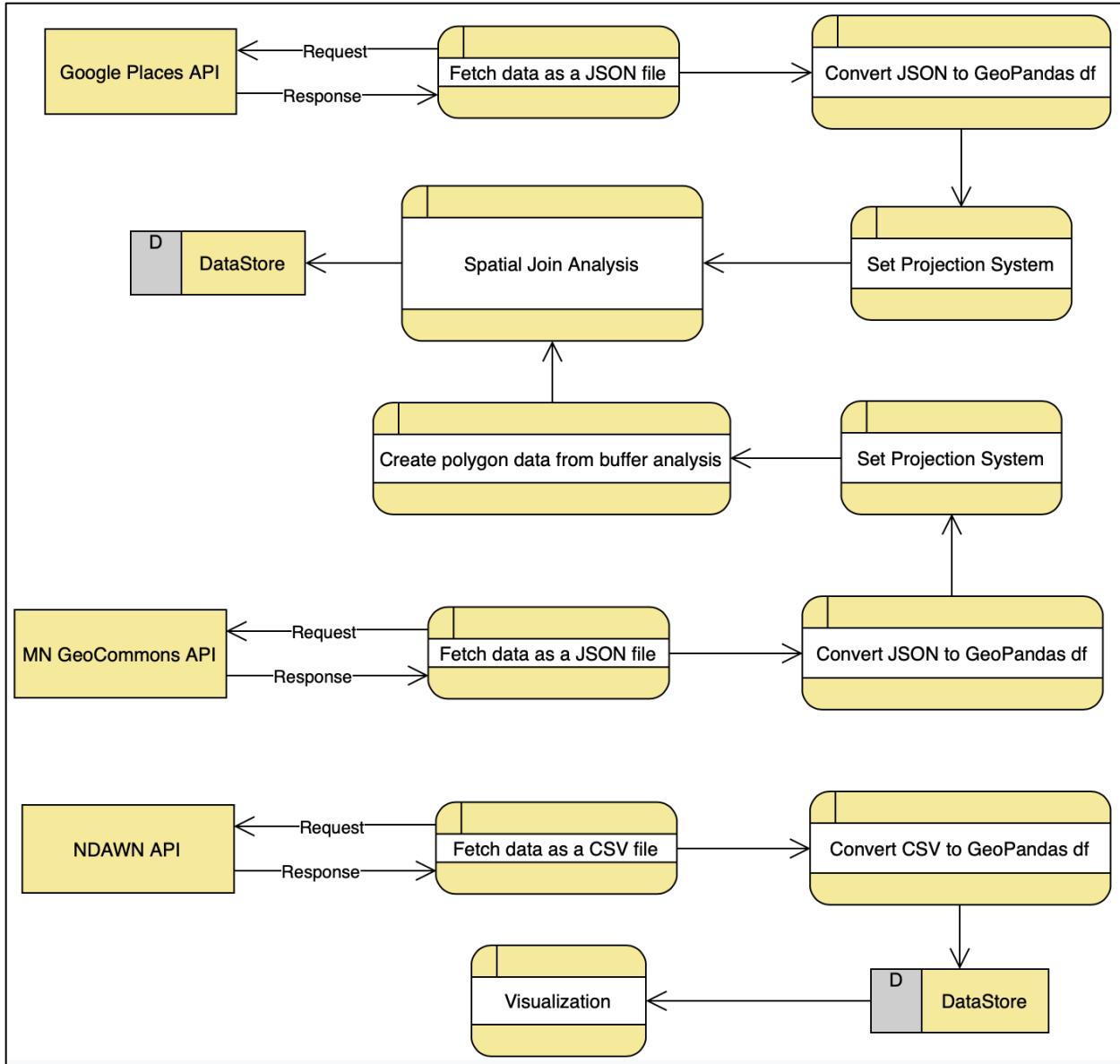


Figure 1. Data flow diagram.

Results

After downloading the data from google places, MN geocommons, and NDAWN, the results are depicted in figures 1 to 8. According to figure 2, the coffee shop data includes parameters like name, rating, and latitude and longitude. This data is depicted in figure 3 using folium as an interactive map. A blue circle of radius 3 km is also shown that limits the study area for this project. Figure 4 shows an instance of the MPCA data that includes parameters like site name, zip code, latitude and longitude. After running a buffer analysis with the radius of 1500m for each point, the points as well as the polygon data resulted from the buffer analysis are visualized in figure 5. Figure 6 and 7 shows the results of spatial join analysis. It contains 32 columns that is the summation of MPCA data (23 columns) and coffee shop data (9 columns). The black markers in figure 7 illustrates the coffeeshops that are in the neighborhood of 5 km and have had

MPCA sites in the neighborhood of 1.5 km of them. This analysis resulted in 8 coffee shops of interest suitable for me to work on my dissertation.

| | name | business_status | open_now | rating | address | lat | lng | geometry |
|---|--------------------------------|-----------------|----------|--------|---------------------------------|-----------|------------|----------------------------|
| 0 | Alle Aamin Coffee Shop | OPERATIONAL | True | 4.5 | 609 Cedar Ave, Minneapolis | 44.967374 | -93.247063 | POINT (-93.24706 44.96737) |
| 1 | Midway Coffee and Tea | OPERATIONAL | True | 5.0 | 1607 University Ave W, St Paul | 44.956069 | -93.168279 | POINT (-93.16828 44.95607) |
| 2 | Lori's Coffee House | OPERATIONAL | False | 4.6 | 1441 Cleveland Ave N, St Paul | 44.984514 | -93.187575 | POINT (-93.18757 44.98451) |
| 3 | Starbucks | OPERATIONAL | False | 3.8 | 1536 Hewitt Ave, St Paul | 44.964455 | -93.166626 | POINT (-93.16663 44.96446) |
| 4 | Up Cafe and Up Coffee Roasters | OPERATIONAL | False | 4.8 | 1901 Traffic St NE, Minneapolis | 44.992616 | -93.225131 | POINT (-93.22513 44.99262) |

Figure 2: The head of the coffee shop data downloaded from google places API

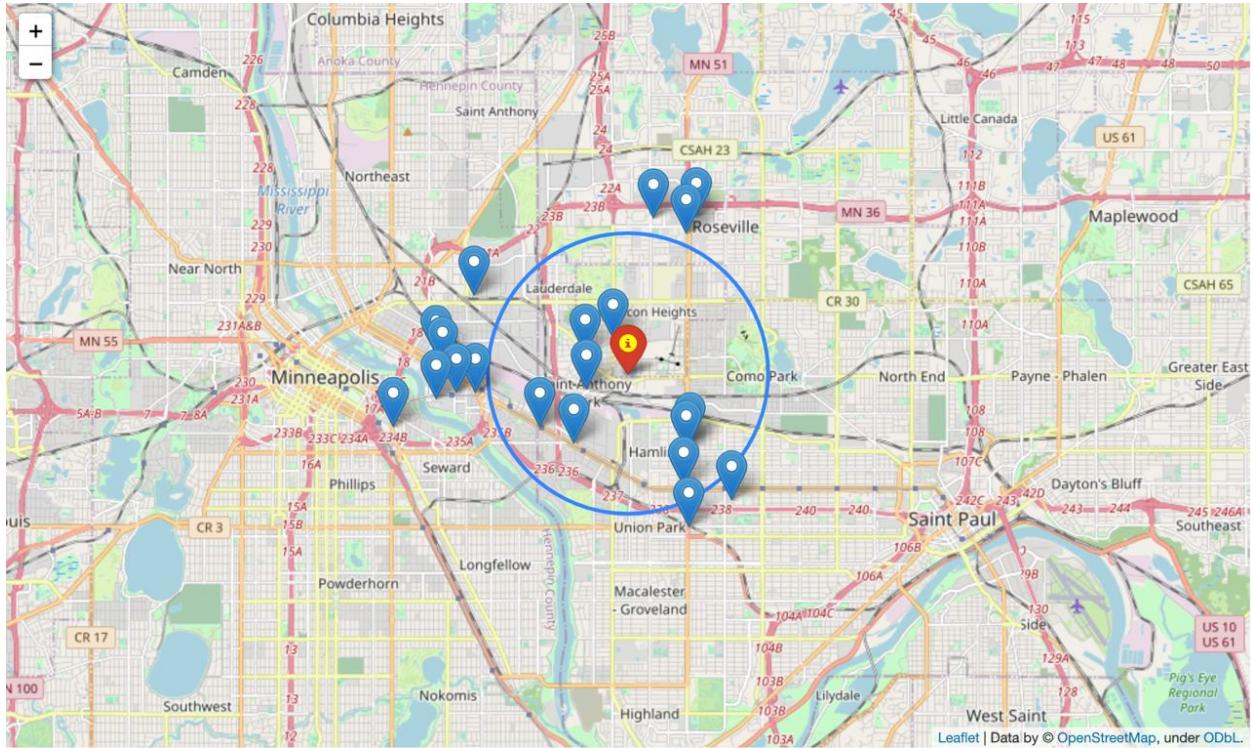


Figure 3: All the coffee shops in 5 km neighborhood of my house. The blue circle shows a 3 km neighborhood of my place.

| zipCode | watershedUrl | huc8 | lat | siteName | coordinateMethod | legislativeDistrict | watershedName | county |
|---------|---|----------|-------------|---|-----------------------------|---------------------|---------------------------------|--------|
| 9439 | 55120 http://www.pca.state.mn.us/index.php/water/wat... | 07020012 | 44.867066 | BRT Mendota Heights Trailhead | Digitized - MPCA online map | 52B | Lower Minnesota River | I |
| 10380 | 55106 http://www.pca.state.mn.us/index.php/water/wat... | 07010206 | 44.963278 | Beacon Bluff Parcel 2 | Digitized - MPCA online map | 67A | Mississippi River - Twin Cities | R |
| 13484 | 55111 http://www.pca.state.mn.us/index.php/water/wat... | 07020012 | 44.88351964 | Boy Scouts of America Northern Star Council Le... | Zip Code Centroid | 63B | Lower Minnesota River | He |
| 23734 | 55105 http://www.pca.state.mn.us/index.php/water/wat... | 07010206 | 44.94021 | Commercial Building | Digitized - MPCA online map | 64A | Mississippi River - Twin Cities | R |
| 23888 | 55103 http://www.pca.state.mn.us/index.php/water/wat... | 07010206 | 44.9847 | Como Golf Course Stormwater BMPs | Digitized - MPCA online map | 66B | Mississippi River - Twin Cities | R |

5 rows × 23 columns

Figure 4: The head of the MPCA data downloaded from the MN Geocommons' API.

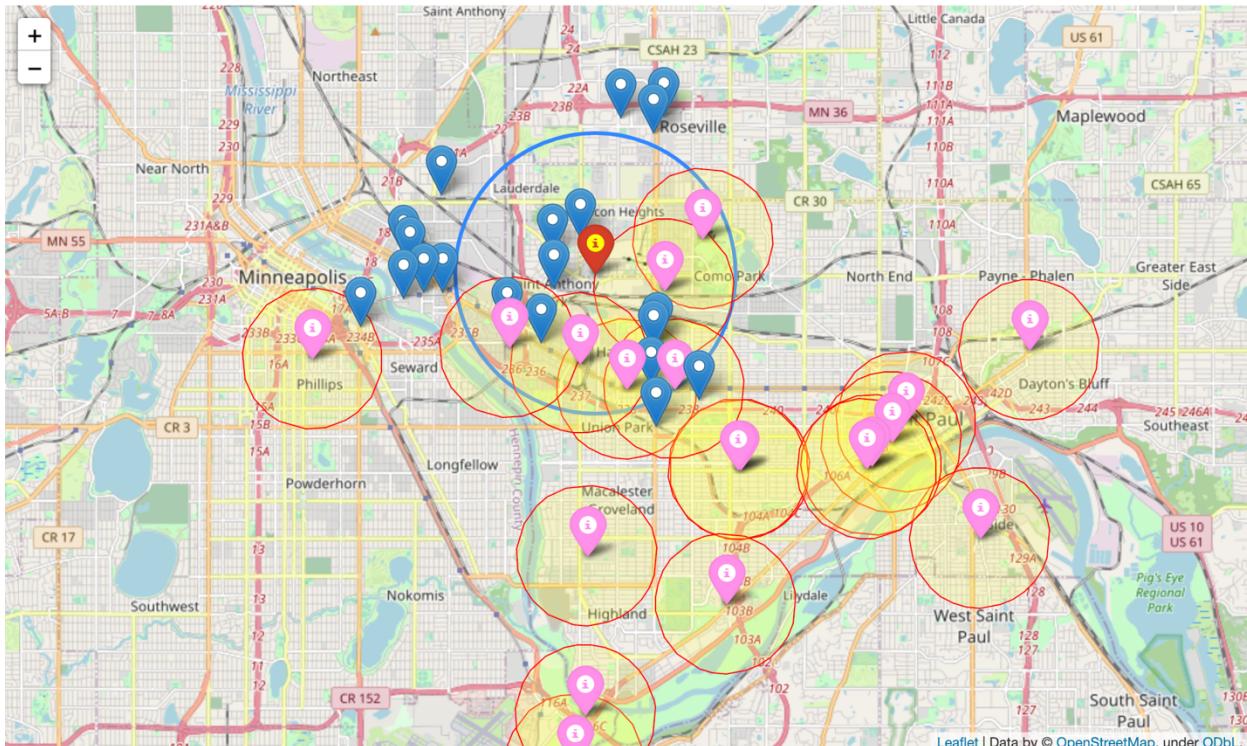


Figure 5: The visualization of the MPCA sites and coffee shops in my neighborhood. The circles around pink markers show a neighborhood of 1.5 km for each MPCA site.

| | name | business_status | open_now | rating | address | lat_left | lng | geometry_left | index_right | zipCode | ... | stateCode | institutionalControl |
|----|------------------------|-----------------|----------|--------|--------------------------------|-----------|------------|----------------------------|-------------|---------|-----|-----------|----------------------|
| 0 | Alle Aamin Coffee Shop | OPERATIONAL | True | 4.5 | 609 Cedar Ave, Minneapolis | 44.967374 | -93.247063 | POINT (-93.24706 44.96737) | 92256 | 55404 | ... | MN | N |
| 1 | Midway Coffee and Tea | OPERATIONAL | True | 5.0 | 1607 University Ave W, St Paul | 44.956069 | -93.168279 | POINT (-93.16828 44.95607) | 31657 | 55104 | ... | MN | N |
| 3 | Starbucks | OPERATIONAL | False | 3.8 | 1536 Hewitt Ave, St Paul | 44.964455 | -93.166626 | POINT (-93.16663 44.96446) | 31657 | 55104 | ... | MN | N |
| 11 | Starbucks | OPERATIONAL | False | 3.8 | 234 Snelling Ave N, St Paul | 44.948251 | -93.166686 | POINT (-93.16669 44.94825) | 31657 | 55104 | ... | MN | N |
| 19 | Ginkgo Coffeehouse | OPERATIONAL | False | 4.4 | 721 Snelling Ave N, St Paul | 44.963016 | -93.167383 | POINT (-93.16738 44.96302) | 31657 | 55104 | ... | MN | N |

5 rows × 32 columns

Figure 6: The head of the table resulted from spatial join analysis. It contains 32 columns that is the summation of MPCA data (23 columns) and coffee shop data (9 columns).

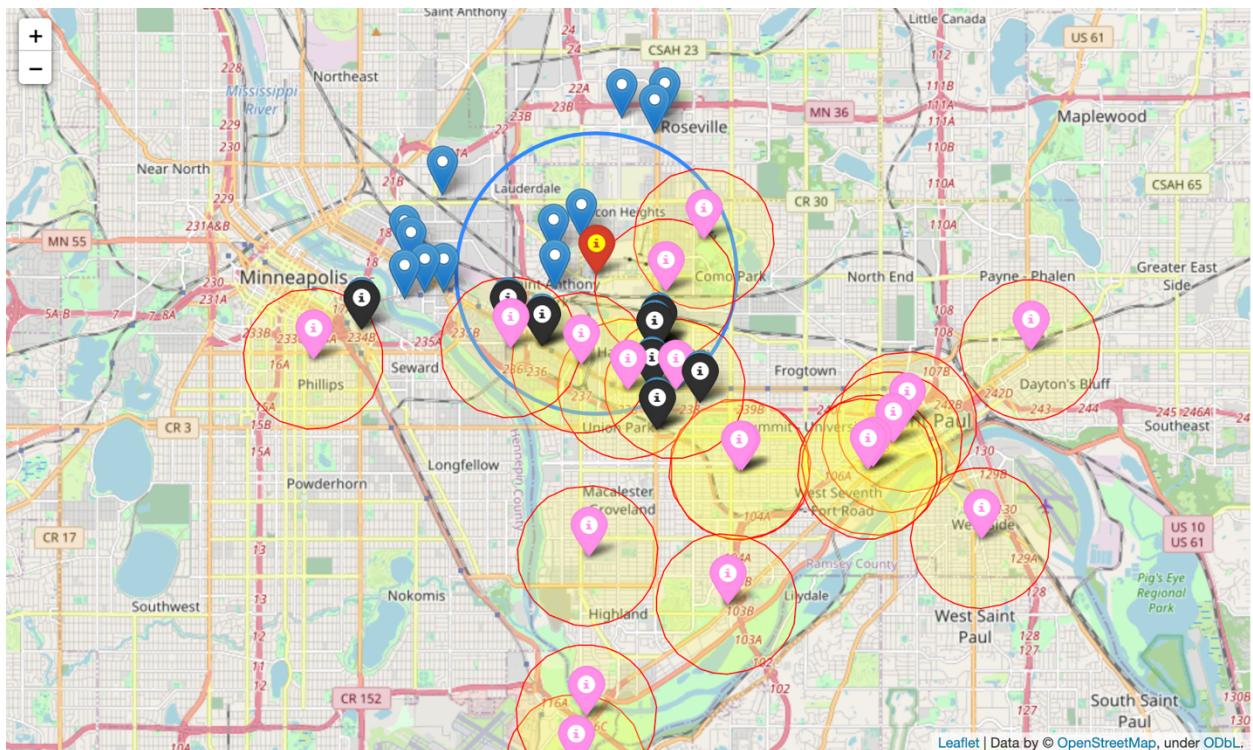


Figure 7: The visualization of data resulted from spatial join analysis on coffee shops and MPCA sites data.

| | Station Name | Latitude_deg | Longitude_deg | Elevation_ft | Year_Unnamed_4_level_1 | Month_Unnamed_5_level_1 | Day_Unnamed_6_level_1 | Max Temp_Degrees_F | Max Temp_Degrees_F | Min Temp_Degrees_F | Max ... |
|---|------------------------|--------------|---------------|--------------|------------------------|-------------------------|-----------------------|------------------------|------------------------|-------------------------|-------------------------|
| | Name_Unnamed_0_level_1 | Latitude_deg | Longitude_deg | Elevation_ft | Year_Unnamed_4_level_1 | Month_Unnamed_5_level_1 | Day_Unnamed_6_level_1 | Flag_Unnamed_8_level_1 | Flag_Unnamed_8_level_1 | Flag_Unnamed_20_level_1 | Flag_Unnamed_20_level_1 |
| 0 | Eldred | 47.68800 | -96.822000 | 861 | 2021 | 9 | 23 | 77.324 | NaN | 49.874 | ... |
| 1 | Perley | 47.17900 | -96.680000 | 895 | 2021 | 9 | 23 | 76.460 | NaN | 48.776 | ... |
| 2 | Humboldt | 48.88400 | -97.150000 | 798 | 2021 | 9 | 23 | 79.376 | NaN | 48.578 | ... |
| 3 | Stephen | 48.45675 | -96.853953 | 1072 | 2021 | 9 | 23 | 77.684 | NaN | 51.422 | ... |
| 4 | Warren | 48.13700 | -96.839000 | 854 | 2021 | 9 | 23 | 76.802 | NaN | 49.532 | ... |

Figure 8: The head of the NDAWN meteorological data.

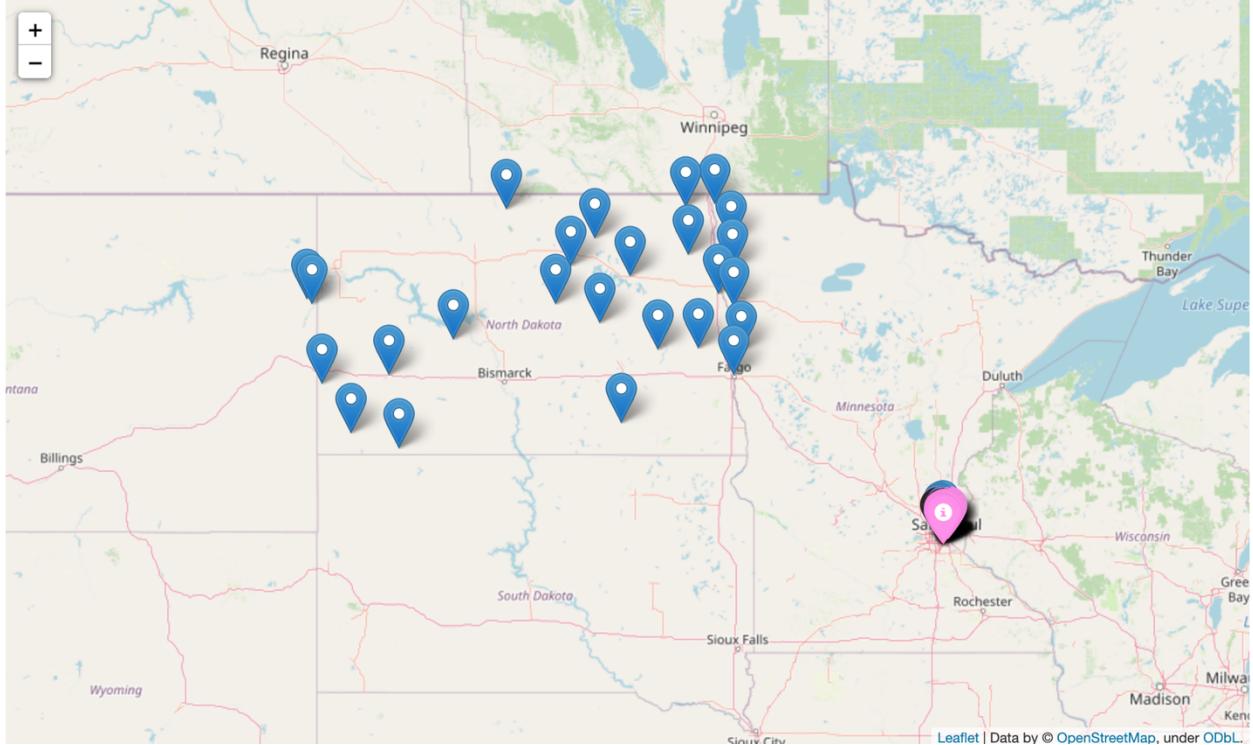


Figure 9: The visualization of all raw and processed data on the same map. This map is interactive in the notebook with pop up messages and mouse hover messages.

Results Verification

One approach for evaluating and verifying the results is visualization. The interactive folium map depicted in figure 9 can be a solid proof for the precision of the results. Comparing the figures and testing the information in the pop-up messages resulted from mouse-over and click events can verify that the results are correct. Since the transformation of all datasets to a specific projection system have been done correctly, the results appear on the map correctly.

Discussion and Conclusion

API for data collection

In this project I used API calls for three data warehouses namely Google Places, Minnesota Geospatial Commons, and NDAWN. This project was a gateway to the world of google cloud platform as well as other online data warehouses. It showed that how many resources are available to download spatial data and services through an automated pipeline. The main challenge in this matter is the parameters. The API documentations for API enabled data warehouses are not always available. Finding the available parameters for each dataset was the main challenge for me to deal with. Also, each data provider has their own data format and structure. Unlike google places and MN geospatial commons, the NDAWN API calls return the data in CSV format. It required me to save it on my machine and then read it again and manipulate it. These are the challenges I see when using API calls.

References

1. Minnesota IT Services, Geospatial [accessed on September 29, 2021]
<https://mn.gov/mnit/government/services/exec/saas/geospatial.jsp>
2. Documentation for "/api/v1/wimn/sites" [accessed on September 29, 2021]
<http://services.pca.state.mn.us/api/doc/v1/wimn/sites>
3. NDAWN GitHub repository [accessed on September 29, 2021]
https://github.com/gettecr/get_us_weather
4. Places API [accessed on September 29, 2021]
<https://developers.google.com/maps/documentation/places/web-service/search-nearby>

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 | 26 |
| 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 | 20 |
| 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 | 94 |