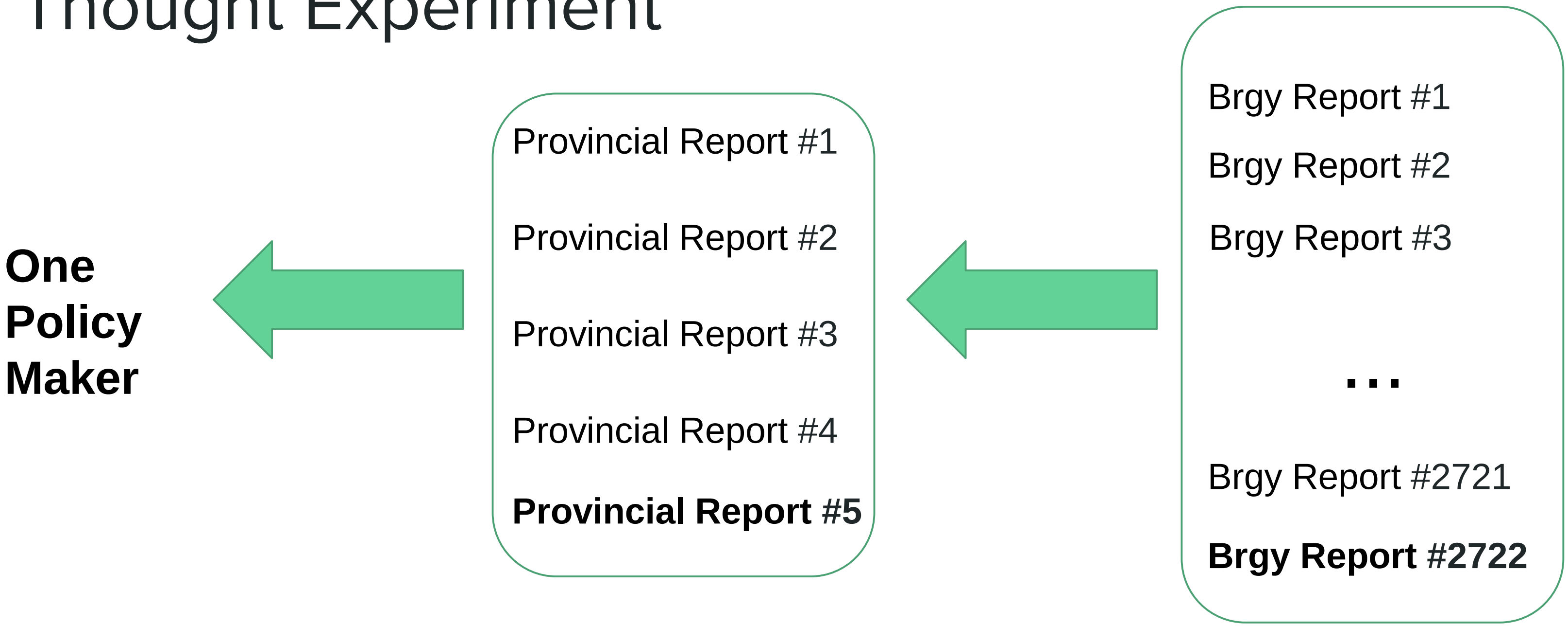
Spatial Analysis & Interactive Visualization for BARMM

Aldrin Lambon

A Thought Experiment



Let’s assume that all the reports can be available in 1 year, then let’s say a policy maker needs another 6 months to consolidate everything before making a decision

This is a long and slow process...

Problem Statement

“Managing these political, spatial, and fiscal changes as the Bangsamoro Transition Authority (BTA) prepares the building-blocks for the Bangsamoro Parliament in 2022 will be challenging, to say the least, given competing **parallel priorities affecting the roughly 4.64 million residents** of the area. All of these will require **thousands of simultaneous micro-decisions** to be made at the same time, as well as the **evidence required to make these decisions**.”

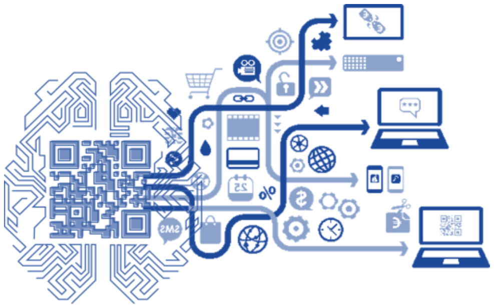
Excerpt from [Open Data for the Bangsamoro Transition Practical notes on using geospatial and statistical data for decision-making]

Solution Overview

Database Architecture

Interactive Dashboard

Recommender System



Doing all of this needs a complete data science team from data engineers to domain experts. As a Proof of Concept (POC), I will be presenting to you a part of this solution, an Interactive Dashboard that utilizes both spatial and tabular data

Objective

1. To identify population hot spots and cold spots in the Bangsamoro Autonomous Region Provinces and Barangays
2. To utilized both tabular and spatial data into an interactive dashboard to maximize insight and functionality for policy makers
3. To provide an overview of the features that can be related to sustainable development goals (SDGs)

Data Coverage

Spatial Data:

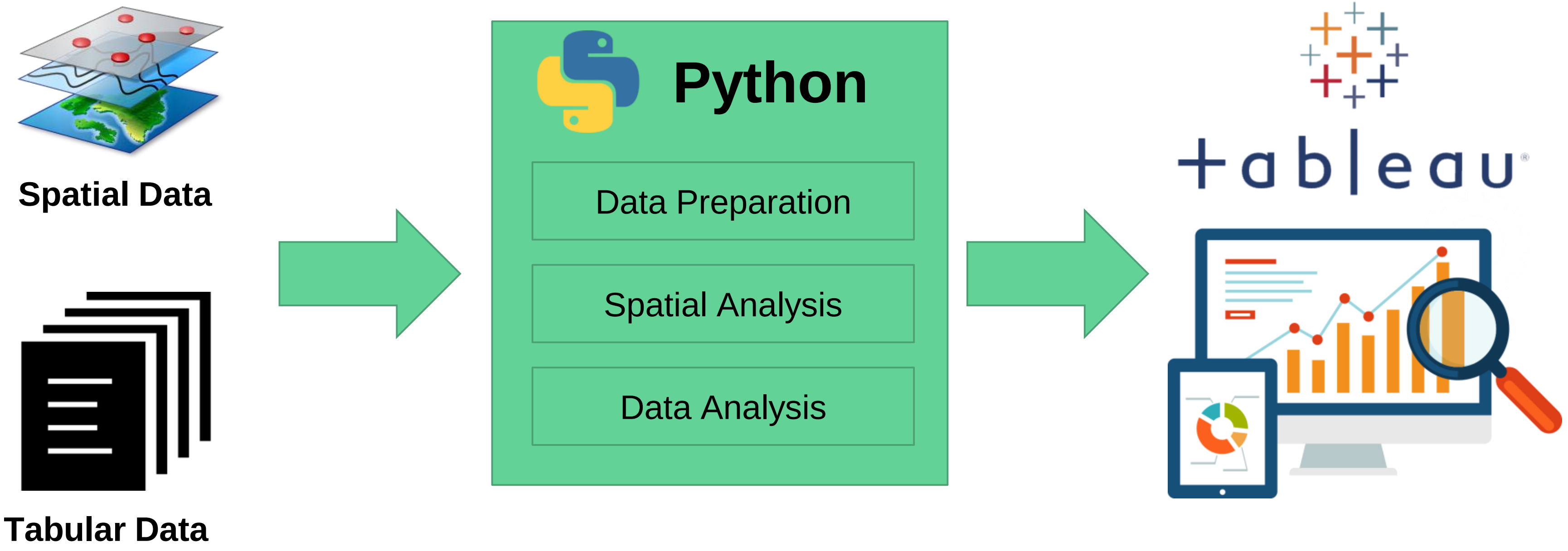
Administrative Boundaries ARMM Barangays (PSA, 2016), Administrative Boundaries ARMM Provinces (PSA, 2016)

Tabular Data:

Population, Number of Schools, Water Supply, Fuel for Lighting, Worker Occupation and Construction Material

Source: https://www.openbangsamoro.com/

Methodology



Spatial Analysis

Univariate Spatial Clustering of Population Totals at the Brgy Level

Spatial Weights

Queen-based Contiguity
- All barangays that share the same borders are considered neighbors

Spatial Statistics

Moran's I
- measures spatial autocorrelation based on both feature locations and feature values simultaneously

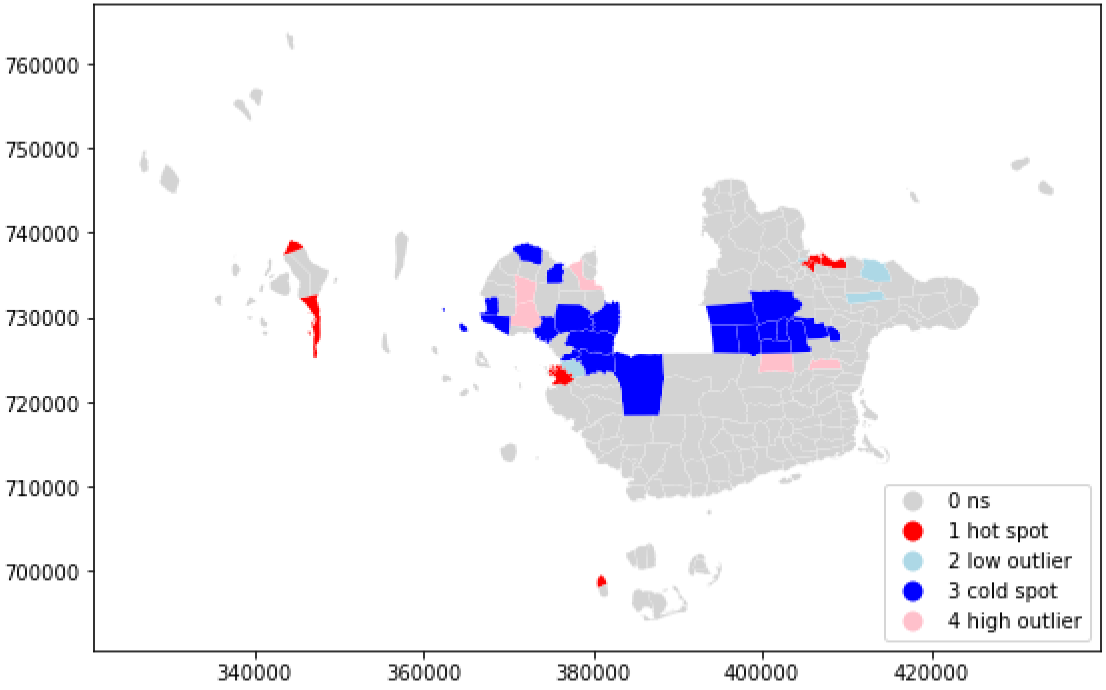
Spatial Clusters

Local Indicators of Spatial Association (LISA)
- can identify spatial dependency (hot spots, cold spots, and spatial outliers) in a given locality

Appendix

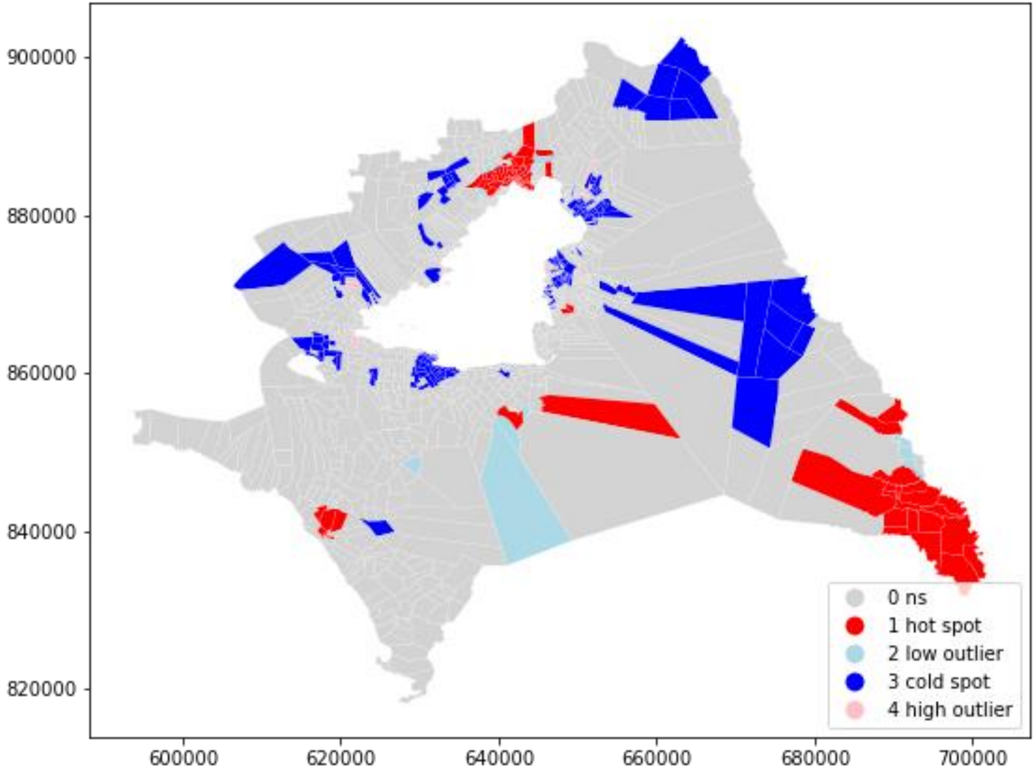
Basilan: Spatial Clusters

Cluster	# Brgys
Hotspot	13
Coldspot	21
High Outlier	5
Low Outlier	4



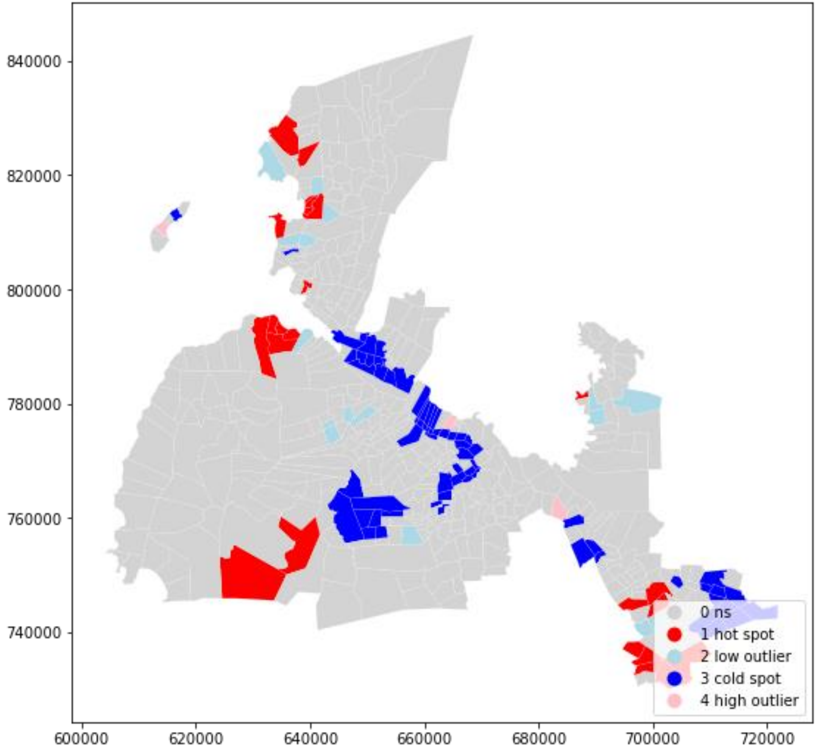
Lanao Del Sur: Spatial Clusters

Cluster	# Brgys
Hotspot	93
Coldspot	182
High Outlier	8
Low Outlier	16



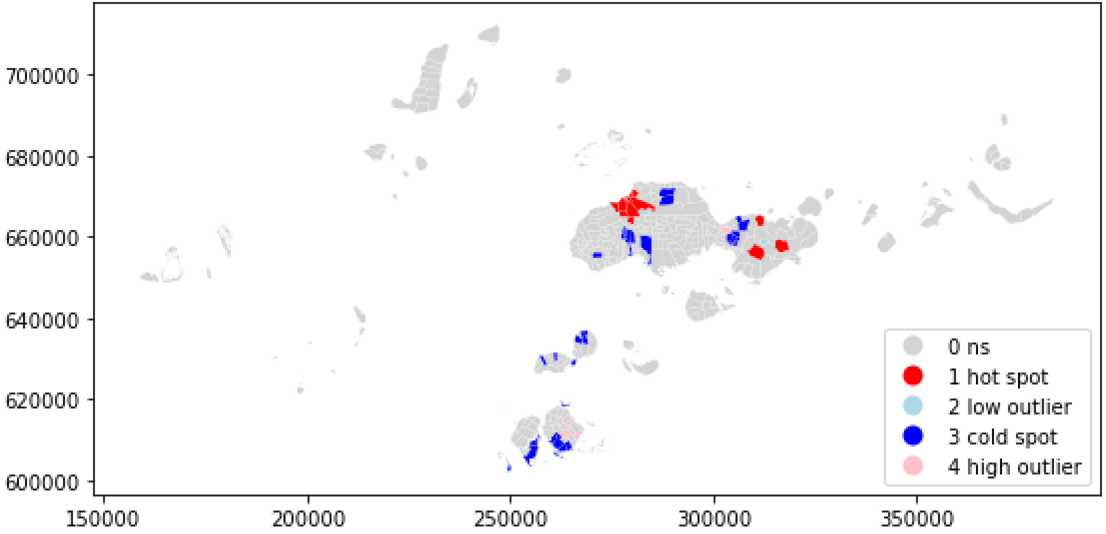
Maguindanao: Spatial Clusters

Cluster	# Brgys
Hotspot	25
Coldspot	69
High Outlier	3
Low Outlier	17



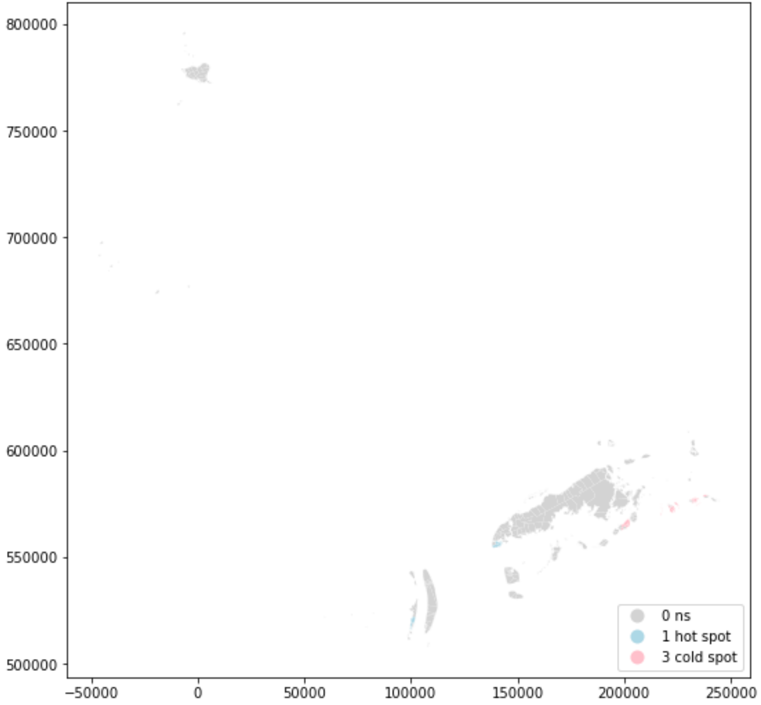
Sulu: Spatial Clusters

Cluster	# Brgys
Hotspot	18
Coldspot	43
High Outlier	5
Low Outlier	2



Basilan: Spatial Clusters

Cluster	# Brgys
Hotspot	12
Coldspot	28
High Outlier	0
Low Outlier	0

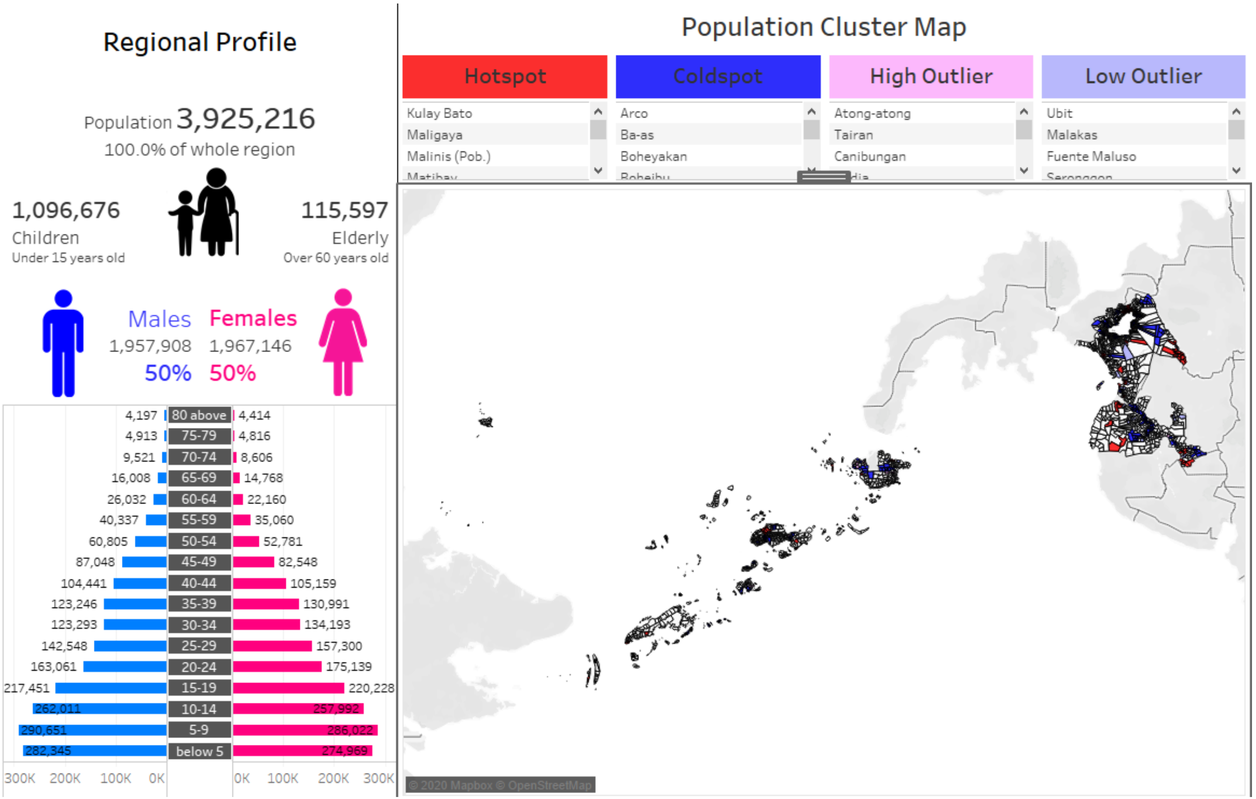


Interactive Visualization

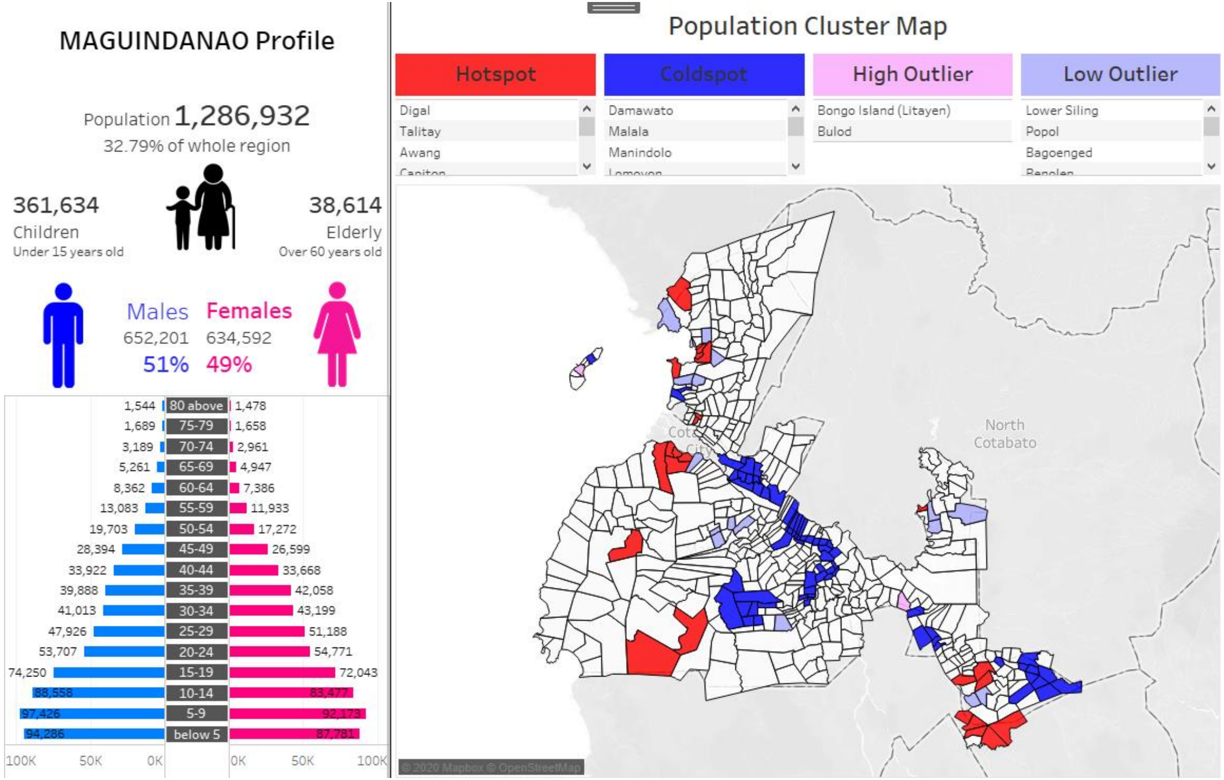
DEMO

https://tabsoft.co/2FnC5Wn

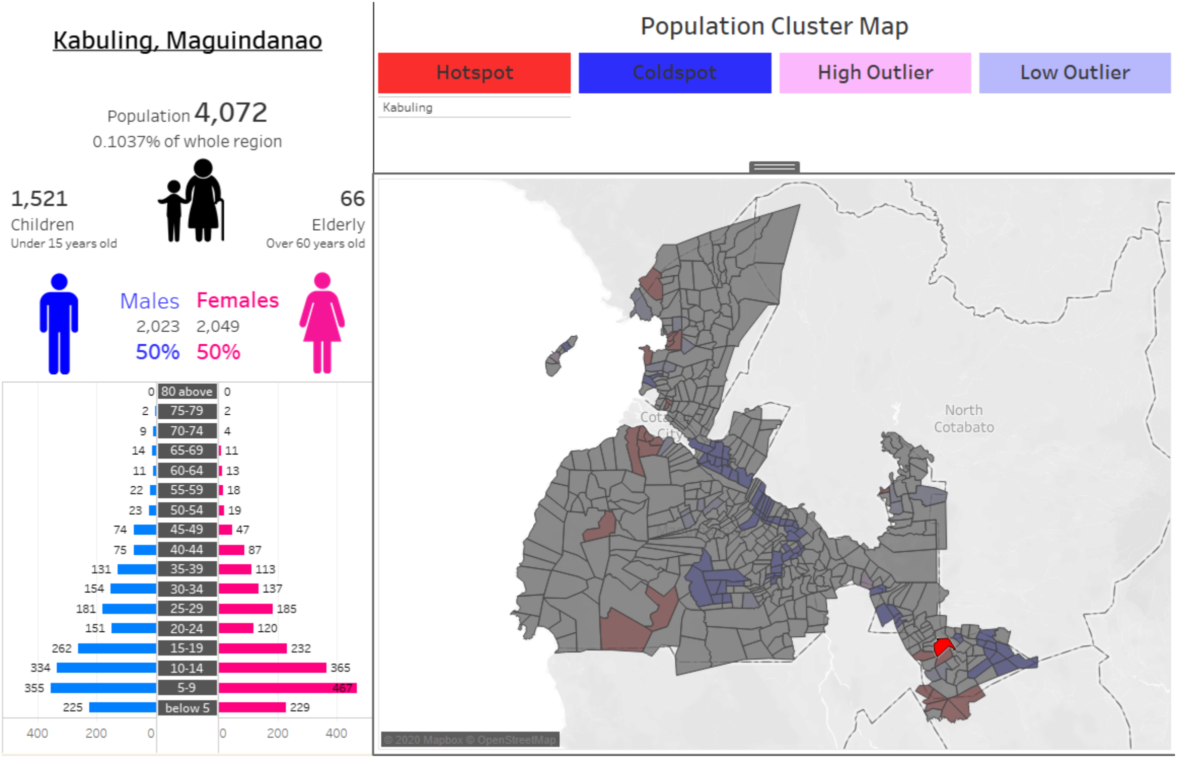
Contains Population Structure along with the Population Cluster Map



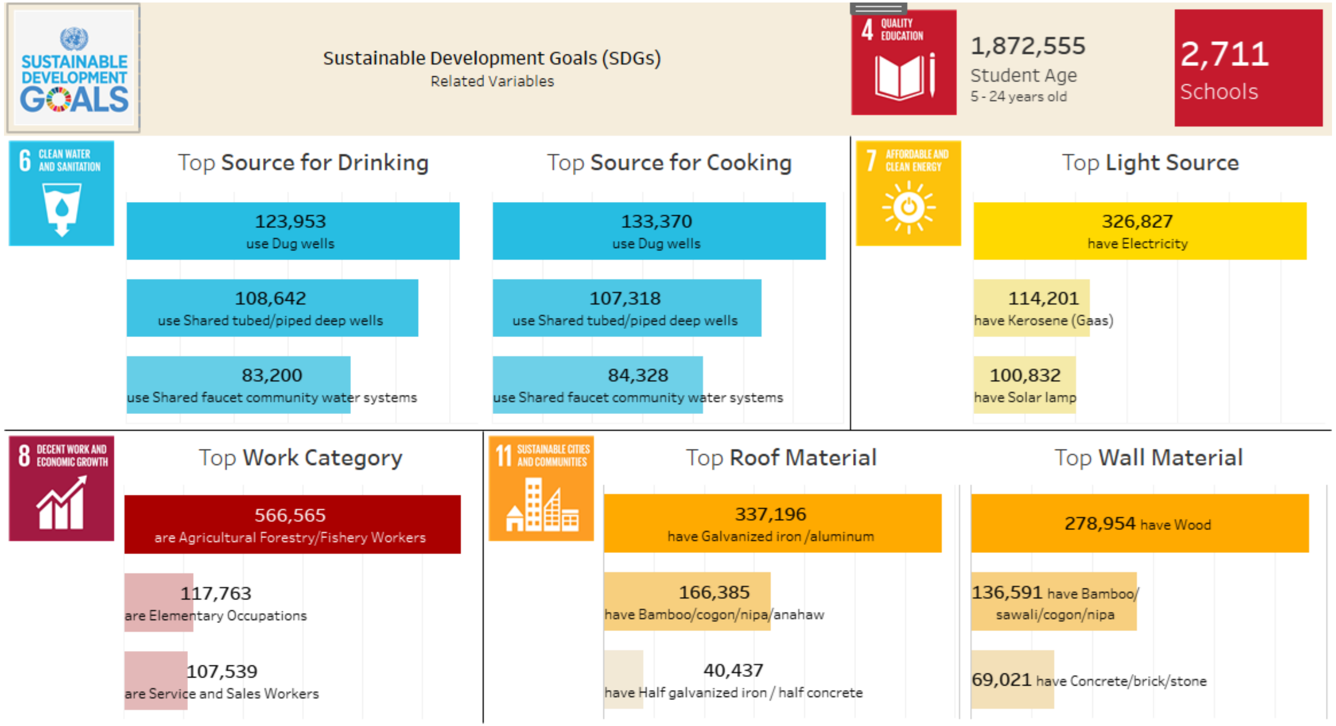
You can filter to a specific province



You can filter also filter to a specific barangay



Along with the provincial and barangay filter, you can also see below the dashboard SDGs related variables the a selected location



Conclusion

- Doing spatial analysis on the population data will aid policy makers identify areas that should be priority zones for new policies or projects
- Integrating the results of spatial analysis to an interactive tableau dashboard, the policy maker can explore multiple data sources in a digestible and flexible manner. From viewing at the regional to filtering down to the barangay level
- Due to the nature of the visualization software used, the policy maker can even add new data sources and add new charts to further improve the analysis

Moving Forward

Database Architecture

- It is essential to have the right database architecture in place for faster access of data and the ability to expand to other data related projects.
- An application can also be develop to upload data directly to the database, this means that policy makers can easily get .

Needed Collaboration:

- Data Engineers
- Software Engineers

Moving Forward

Recommender System

- Policy makers can speed up their decision making process by utilizing a recommender system that can analyze previously implemented policies and correlate it to the characteristics of a specific location

Needed Collaboration:

- Data Scientist
- Policy Makers (Domain Experts)

Spatial Analysis & Interactive Visualization Project

Aldrin Lambon

Appendix: Global Moran's I

The Moran's I statistic for spatial autocorrelation is given as:

$$I = \frac{n}{S_0} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,j} z_i z_j}{\sum_{i=1}^{n} z_i^2} \qquad (1)$$

where z_i is the deviation of an attribute for feature i from its mean $(x_i - \bar{X})$, $w_{i,j}$ is the spatial weight between feature i and j, n is equal to the total number of features, and S_0 is the aggregate of all the spatial weights:

$$S_0 = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,j} \qquad (2)$$

The z_I-score for the statistic is computed as:

$$z_I = \frac{I - \mathrm{E}[I]}{\sqrt{\mathrm{V}[I]}} \qquad (3)$$

where:

$$\mathrm{E}[I] = -1/(n-1) \qquad (4)$$

$$\mathrm{V}[I] = \mathrm{E}[I^2] - \mathrm{E}[I]^2 \qquad (5)$$

https://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/h-how-spatial-autocorrelation-moran-s-i-spatial-st.htm

Appendix: Local Moran's I

The Local Moran's I statistic of spatial association is given as:

$$I_i = \frac{x_i - \bar{X}}{S_i^2} \sum_{j=1, j\neq i}^{n} w_{i,j}(x_j - \bar{X}) \qquad (1)$$

where x_i is an attribute for feature i, $\bar{X}$ is the mean of the corresponding attribute, $w_{i,j}$ is the spatial weight between feature i and j, and:

$$S_i^2 = \frac{\sum_{j=1, j\neq i}^{n} (x_j - \bar{X})^2}{n-1} \qquad (2)$$

with n equating to the total number of features.

The z_{I_i}-score for the statistics are computed as:

$$z_{I_i} = \frac{I_i - \mathrm{E}[I_i]}{\sqrt{\mathrm{V}[I_i]}} \qquad (3)$$

where:

$$\mathrm{E}[I_i] = -\frac{\sum_{j=1, j\neq i}^{n} w_{ij}}{n-1} \qquad (4)$$

$$\mathrm{V}[I_i] = \mathrm{E}[I_i^2] - \mathrm{E}[I_i]^2 \qquad (5)$$

https://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/h-how-cluster-and-outlier-analysis-anselin-local-m.htm