GATEWAY: A GEOSPATIAL ANALYTICS SYSTEM

L.J. Lao [1], B.J. Harder [1]

[1] Cobena Business Analytics and Strategy, Inc., BGC, Taguig, Philippines – leonalao@cobenagroup.com,
benharder@cobenagroup.com

Commission IV

KEYWORDS: geospatial, data analytics, geovisualization, machine learning, web platform, location analysis

ABSTRACT:

This paper presents an overview of the Gateway web platform, a proprietary geospatial analytics system developed by Cobena Business Analytics and Strategy, Inc.[1] The application is intended to serve as a user-friendly and easily-accessible tool for spatial data analysis and visualization geared toward non-technical specialists. Gateway's core functionalities hinge on mapping and data visualization (choropleths and points) alongside traditional scoring methods and built-in machine learning algorithms for area prioritization and site selection. Gateway provides an interactive, cloud-based environment that abstracts and simplifies common location-based analyses. A core strength of the platform is also its heavy localization to the Philippine context through a curated database of market information — with future plans to create local counterparts across SEA — which reduces the need for extensive external market data collection and reconciliation. The paper gives a brief review of the system design and key features of the platform. It also highlights some key applications across industries such as real estate, consumer goods, and retail in informing expansion and distribution strategies, prioritizing resource allocation, and analyzing historical performance against market factors.

1. INTRODUCTION

1.1 Data and Spatial Challenges in the Philippines

The primary challenge of many Philippine organizations is the seeming data scarcity in the age of big data. While there is an abundance of internal data typically collected by each organization, businesses require external market information to make properly informed business decisions.

Many geographic information systems (GIS)—which are typically employed in spatial analyses—necessitate organizations to have or purchase datasets before users can conduct relevant analyses for their business. ESRI offers ESRI Data, a "global database of maps and data,"[2] but only administrative boundaries, average household size, and purchasing power per capita datasets are currently available for the Philippines.[3]

Initiatives such as PhilGIS.org[4], Open Data Philippines[5], Philippine Statistics Authority (PSA)'s OpenStat[6], or the National Mapping and Resource Authority (NAMRIA) Geoportal[7] have addressed some key needs of data collation, but many of these are largely limited in scope and/or require a lot of data transformation to be ready for analytical processing. As such, while a sizable amount of external market information and official statistics exist in the country,

no central database is readily utilized by and available to those in the private sector.

Further, many individuals and organizations encounter issues of data integrity within and across datasets and data sources. There are inconsistent names across datasets (e.g. variations in spelling). While the Philippine Standard Geographic Code (PSGC) is meant to serve as the country's uniform identifier, the coding system also changes across time, leading to non-uniformity. There are also issues of comparability of data across time and regularity in dates (e.g. labor categories between 2014 and 2019 PSA reports, frequent major administrative area changes such as the formation and dissolution of the Negros Island Region).

1.2 Spatial Data Applications in the Philippines

Despite these challenges, the Philippines is not a newcomer in geospatial sciences. While some organizations in the energy industry, real estate and development, utilities, and food industries have utilized GIS platforms, the most extensive and advanced users of geospatial technologies in the country are in the academe and government.

---

[1] Cobena Business Analytics and Strategy, Inc. is a management consultancy and data analytics company based in Bonifacio Global City, Taguig, Philippines.

[2] https://www.esri.com/en-us/arcgis/products/esri-demographics/get-started

[3] https://livingatlas.arcgis.com/

[4] http://philgis.org/

[5] https://data.gov.ph/homepage

[6] http://openstat.psa.gov.ph/

[7] http://www.geoportal.gov.ph/

Two examples of these are Project NOAH[8] (Nationwide Operational Assessment of Hazards) and the DREAM/Phil-LiDAR 1 Program, both of which are partnerships between the University of the Philippines (UP) and the Department of Science and Technology (DOST). Project NOAH provides a portal by which Filipinos can access spatial information on all available natural risks and hazards. It serves as a central data repository of weather, floods, landslides, storm surges, earthquakes, volcanoes, and even dengue information.

The DREAM/Phil-LiDAR 1 Program[9], on the other hand, pioneered the use of LiDAR (Light Detection and Ranging) technology in the country that allowed the rapid mapping of large areas at very high resolutions (1m). The DREAM/Phil-LiDAR 1 Program allowed for the development of high-resolution flood hazard maps covering about 2/3 of the country's river systems.

While both of these projects used advanced geospatial tools and technologies for data gathering and analysis, these require a substantial amount of financial support which most private companies, especially small and medium enterprises (SME's), that are in need of geospatial analysis do not have.

On top of this, specialists are often required to be able to handle geospatial data and transform them in a way that will be understandable to the end-users.

2. DESIGN AND FUNCTIONALITY

Given the challenges of spatial data and limited geospatial applications in the private sector, Gateway was designed and intended to serve as a user-friendly and easily-accessible tool for spatial data analysis and visualization geared toward non-technical specialists.

Gateway's core functionalities hinge on mapping and data visualization (choropleths and points) alongside traditional scoring methods and built-in machine learning algorithms for area prioritization and site selection.

Gateway provides an interactive, cloud-based environment that abstracts and simplifies common location-based analyses. A core strength of the platform is also its heavy localization to the Philippine context through a curated database of market information which reduces the need for extensive external market data collection and reconciliation.

2.1 Solutions to Data-related Challenges

2.1.1 Availability: The Gateway system includes both relational and non-relational databases to store cleaned spatial information from different government agencies, various non-government organizations, and local and global scientific institutions.

2.1.2 Integrity: Collecting data from various sources reaps the additional benefit of being able to validate and flag anomalies in the datasets by comparing each, e.g. Philippine Deposit Insurance Corporation (PDIC) vs Banko Sentral ng Pilipinas (BSP) financial measures, population from the United Nations (UN) vs PSA, road networks of OpenStreetMap (OSM) vs Department of Public Works and Highways (DPWH). Internal processes have also been set in place to remove problematic datasets (e.g. based on % missing data, skew) when running the location analyses. As the datasets are used in analytics runs, other potential anomalies in the data are also surfaced (e.g. upon inspecting the characteristics of the results).

2.1.3 Consistency: Gateway maintains its own identifier protocols that takes into consideration changing PSGC systems, alternate location spellings, and the like across different datasets. Whenever relevant to the dataset, Gateway also maintains interpolated values for missing years so that analyses between datasets are more robust.

2.2 System Design

A key barrier to industry adoption of spatial data analytics is a lack of understanding on the potential applications. This challenge is two-fold: (1) there is a lack of awareness within many businesses on potential applications and (2) there is a skills, competency, and technology gap within the organization to be able to implement and execute such analyses.

To address these, Gateway was designed into modules which provide organizations with a mental model of understanding many of their business needs:

Module	Use Cases
Snapshot	Data exploration and visualization, site and area assessment
Predict	Forecasting
Grade	Ranking, prioritizing, grading
Focus	Site selection

Table. 1. Module names and use cases for Gateway

The above use cases were synthesized after many consulting engagements and surfacing the common trends in enterprise location analytics needs. Standard algorithms and APIs were developed to support and standardize common analytics workflows for each of these scenarios.

---

[8] https://center.noah.up.edu.ph/

[9] https://dream.upd.edu.ph

Module	Algorithms and APIs
Snapshot	Data retrieval and summarization for data layers and points of interest (POIs)
Predict	Linear regression, Bayesian time series (Facebook Prophet)
Grade	Linear combination, Ridge Regression, LASSO Regression, Random Forest Regression
Focus	Linear combination, Logistic Classification

Table 2. Module name and APIs used

Two databases are concurrently maintained: a PostgreSQL instance for structured data and a MongoDB instance for vector data such as points of interest and road networks. The web platform is built using Angular with a Django back-end and hosted on Microsoft Azure, with back-ups in a local cloud service provider.

2.3 Snapshot: Mapping and Visualization

Gateway's Snapshot module is split in the platform into the Filled Maps and Points tabs.

The Filled Maps tab allows users to quickly map choropleths across any of its data layers. Options are available for switching between administrative boundaries, transforming datasets per capita or as percentage change from previous data points, switching color palettes, or setting benchmarks. Users can easily switch between data from different time periods and view summary statistics.
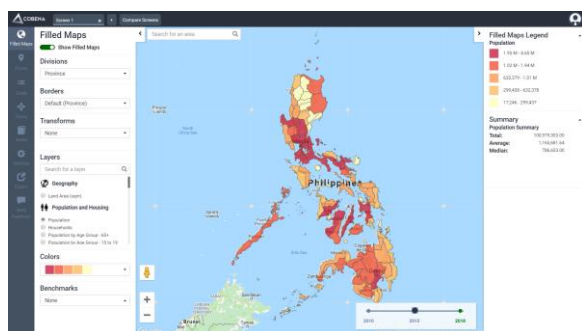

Figure 1. Snapshot: Choropleth display 2015 provincial populations, with legend and summary statistics.

If an organization has a specific focus area, they can also show all sub-areas of a specific place. Figure 2 shows an example of cities and municipalities within the province of Pampanga. The figure also shows how a user can view all key statistics related to an area of interest such as Porac.
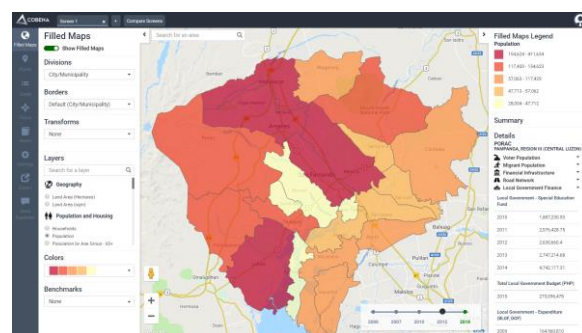

Figure 2. Snapshot: Choropleth of cities and municipalities filtered to a province, with area details summarized in the right sidebar.

The Points tab, meanwhile, allows users to select through POIs, which are hierarchically arranged into categories and subcategories. The POIs automatically cluster or disaggregate in accordance with the current zoom level. Similar to Filled Maps, a legend and summary statistics are also available. Users can also filter to specific areas.
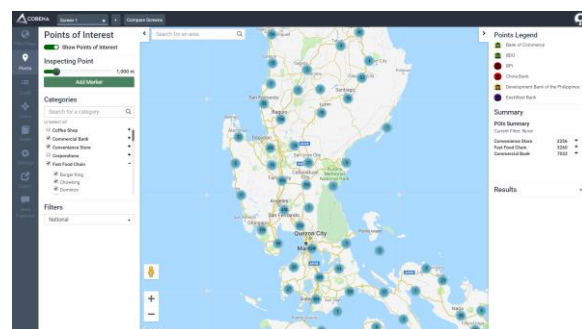

Figure 3. Points of interest clustering according to zoom level, with legend and summary statistics.

Users can also quickly see the profile of a given radius around a site by using the Inspect Point function. The results include a Google Street View image, the nearest POIs, a POI breakdown of surrounding locations, and a list of intersecting areas.
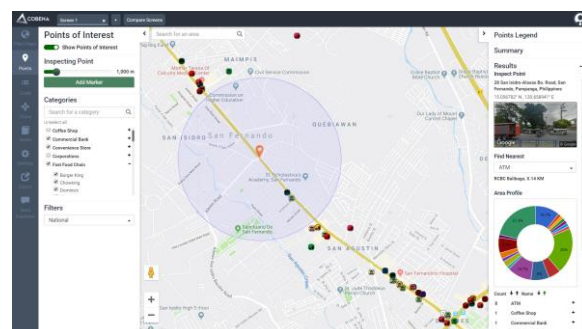

Figure 4. Site inspection for a specified radius, with a profile of the area summarized in the right sidebar.

Users are also able to create multiple screens within Gateway and compare any two screens.
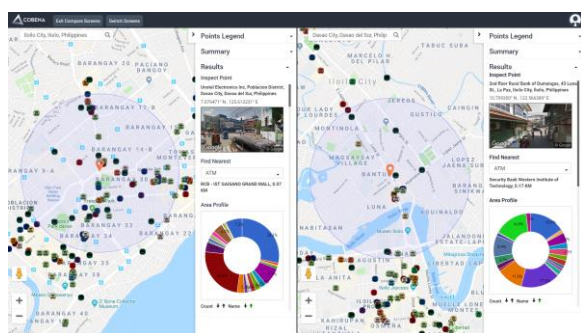
Figure 5. Screens compared side-by-side, typically used when analyzing two locations or results.

2.4 Grade and Focus: Algorithms

Grade and Focus are Gateway's main analysis modules.

Grade is Gateway's area prioritization module. It can be used in two ways: Custom and Automatic. Custom Grade allows users to select and combine various data layers, provide custom weights, and derive a ranking of various areas (region, province, city/municipality, and barangay) based on certain metrics that are important to his/her organization.
2.5
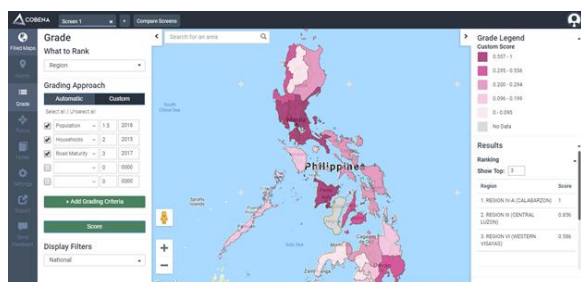
Figure 6. Shows an example of a Custom Grade analysis in Gateway.

Automatic Grade, on the other hand, provides users with the option of uploading their own data to the platform and perform a regression analysis against all data layers from the Gateway database, across the whole country. For example, a company can upload their sales data on a city/municipality level, and Grade will then return a prediction of their sales data across all cities and municipalities in the country. Users can also choose from various preset machine learning models (Random Forest, Ridge, Lasso CV) based on their analysis priority: prediction capability or model interpretability.
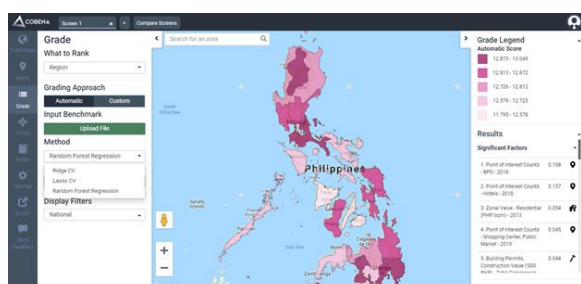

Figure 7. Shows an example of an Automatic Grade analysis using Random Forest Regression

In Figure 7, the Automatic Grade analysis was done using Random Forest Regression on a Regional level. Aside from returning a predicted score for each of the regions in the country, Automatic Grade also returns the list of Significant Factors or important variables determined from the regression analysis.

Focus, on the other hand, is Gateway's Site Selection module. Similar to Grade, it can also be used in two ways: Custom and Automatic Focus.

Custom Focus requires the users to select Supply and Demand Generators. Supply Generators are POIs that provide a similar service or product. For example, if a restaurant is looking for new sites for expansion, Supply Generators would be other competitors and fast food chains. Demand Generators, on the other hand, are POIs that attract the market to an area. Examples of these are transport terminals, schools, malls, etc.
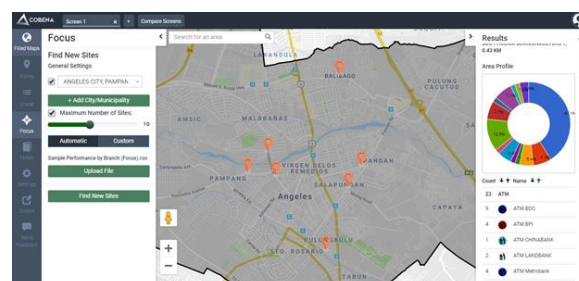

Figure 8. Shows an example of a Custom Focus analysis in Angeles City, Pampanga along with a Site Inspection report on the right sidebar.

3. USE CASES

Gateway has been used across industries such as real estate, logistics, consumer goods, food, and retail in informing expansion and distribution strategies, prioritizing resource allocation, and analyzing historical performance against market factors.

3.1 Site Assessment

Gateway allows users to export a PowerPoint report containing the necessary information about an area. This specific example shows data from Aborlan, Palawan.

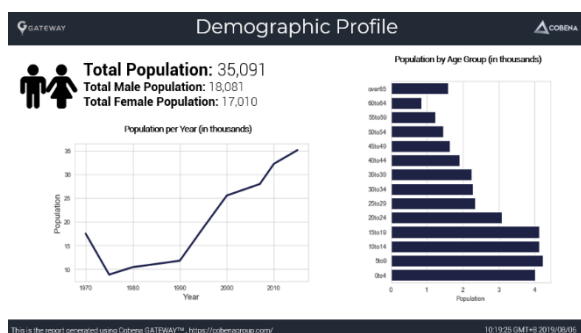Figure 9. Shows the 1st/Homescreen - Quick Statistics of the Gateway Area Report.

Figure 10. Shows the Demographic Profile screen of the Gateway Area Report.
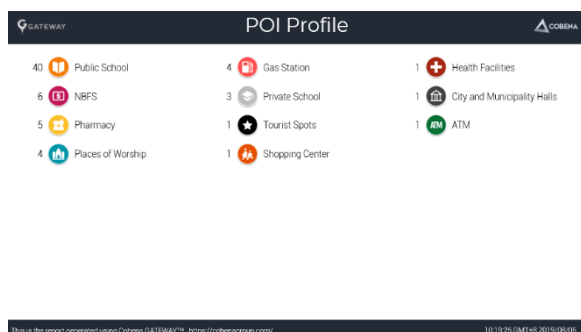

Figure 11. Shows the POI Profile screen of the Gateway Area Report.

Early adopters of the platform mainly used Gateway as a site assessment tool to profile their locations/branches or those of their competitors'.
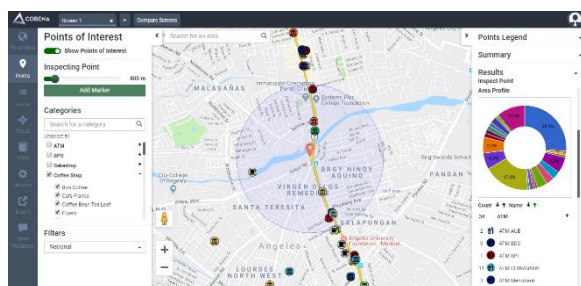

Figure 12. Shows a sample site inspection of Gateway

3.2 Area Prioritization

A certain financial institution wanted to identify which city/municipality in Pampanga would be the "ripest" for them to enter. Using Custom Grade analysis, 4 data layers (Population, Count of Micro, Small, and Medium Enterprises, Total Loans in Php, and Total Savings Deposit) were combined and weighted to identify the top areas for expansion.
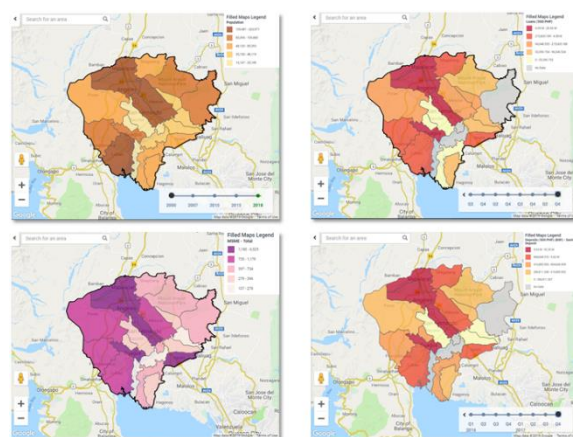

Figure 13. Shows choropleths of the 4 data layers: Population (top left), Loans (top right), MSME Counts (bottom left), and Savings Deposit (bottom right)
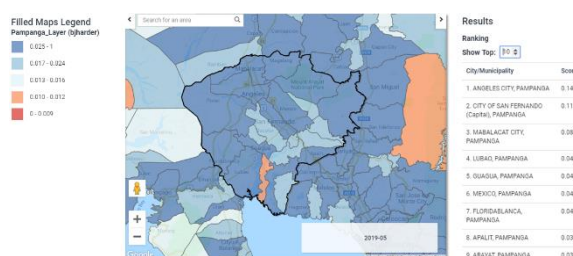

Figure 14. Shows the result of the custom Grade analysis of Pampanga.

Results showed that Angeles, San Fernando, and Mabalacat were the top 3 areas to consider based on the 4 metrics that were used to rank the cities/municipalities in Pampanga.

3.3 Site Selection

3.3.1 Using Custom Focus, a nationwide banking institution identified potential locations for their new branches specifically catering to Filipinos that avail of the Conditional Cash Transfer Program.

3.3.2 Figures 14 and 15 show two examples of the located sites and their area profiles.
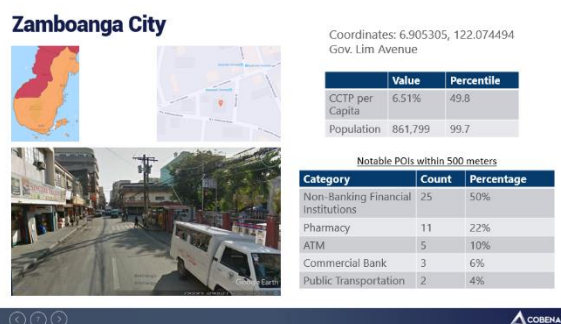
3.3.3

Figure 15. Potential site for a new bank branch in Zamboanga City.

Figure 16. Potential site for a new bank branch in Iligan City.

4. FINDINGS

While there has been extensive use of geospatial analytics in the academe and the public sector, there remains a need to cascade location analytics as a practice within the Philippine private sector, particularly to businesses who many not have the necessary resources (people, investment) to avail of more traditional geospatial analytics services or platforms.

A good location analytics solution in developing countries, particularly those who struggle with the adoption of digital technologies within the industry sector, necessitates that any system must address basic location analytics needs before developing solutions catered to more advanced requirements. Depending on the sophistication of location analytics needs, the models and algorithms need not be very complex to provide actionable solutions to organizations. Further, the structuring and modeling of common industry use cases develops not only confidence in the system, but also lays the necessary foundation for adoption itself by establishing a common framework.

Both the technical and business design of the Gateway system allows for the flexibility required to facilitate industry adoption specifically within the Philippine context, as it addresses critical gaps across the spectrum of needs. However, while the current design and data provides a good starting point, more data is needed to enrich the analyses produced — having laid the groundwork, layering in spatial big data with more velocity becomes an opportunity area (e.g. footfall, geosocial data). While there is a need for more data, of note is also the need for more relevant data — from a computational perspective, many official statistics are redundant (i.e. they behave similarly or explain the same thing, and don't add much to the model).

Aside from improving its underlying data, further work on Gateway will focus on those functionalities which have proven most critical for organizations, alongside the development of more advanced functionality to cater to users further in their location analytics maturity. This encompasses: (a) more efficient visualization and rendering capabilities, (b) handling of more varied spatial data formats, and (c) incorporating more spatial algorithmic techniques

and greater transparency in the machine learning model evaluation metrics.

5. REFERENCES

"Esri Data: Get Started: Living Atlas." Esri. Accessed August 23, 2019. https://www.esri.com/en-us/arcgis/products/esri-demographics/get-started.

Geoportal Philippines. National Mapping and Resource Information Authority. Accessed August 24, 2019. http://www.geoportal.gov.ph/.

"Living Atlas of the World." ArcGIS. Accessed August 23, 2019. https://livingatlas.arcgis.com/.

"Open Data Philippines." Data.gov.ph. Department of Information and Communications Technology. Accessed August 23, 2019. https://data.gov.ph/.

PhilGIS. PhilGIS. Accessed August 23, 2019. http://philgis.org/.

PSA OpenSTAT. Philippine Statistics Authority. Accessed August 23, 2019. http://openstat.psa.gov.ph/.

UP DREAM Program. UP DREAM/PHIL-LiDAR 1. Accessed August 24, 2019. https://dream.upd.edu.ph/.

UP NOAH Center. National Operational Assessment of Hazards, May 10, 2018. https://center.noah.up.edu.ph/.

Revised August 2019