One Acre Fund Written Assessment

Assignment 1:
The Purpose of the assignment is to predict geospatially the likelihood of the presence of banana across Rwanda. The goal is to reveal insight and provide recommendations from data available to stakeholders about the spatial distribution of banana crop across Rwanda, for them to make informed decision.

***Note*:** Some of the images used here are from my other project, so that I can express the context.

Project Layout:

To make a geospatial prediction, I will go through couple of phases to reach the goal of project.

As I do always approach any data science or data project, I use 5 phases/sections to approach a project. Those sections are business problem, data, methodology, analysis, results and conclusion

- *Business problem*: In this section, I provide the high-level overview of the whole project, purpose and tools to conduct the research.
- *Data*: this section is about illustrating the data needed before starting the project and where to retrieve them, this section include a subsection called the data collection, where I illustrated the process I used to get the data.
- *Methodology*: which shows the methods to be used to achieve the goal.
- *Analysis*: In this section, I go deep through my data, checking flaws in data, answer some questions and performing the advanced methods to the data.
- *Results and Discussion*: in this section, I highlight my findings in clear and understanding way, to stakeholders.

Project Implementation:

1. **Data Phase**

After understanding the business problem, I will proceed to data phase where I will gather the data needed for the project. In spatial analysis and prediction, we use geographic data to predict, learn and manage phenomena that affect the earth and its inhabitants [1].

At this stage, I will highlight some data that I will need:

- Shapefile containing information about forests, wilderness areas, and other lands in Rwanda.
- Data about weather in provinces of Rwanda.
- It will be a good idea to use high-resolution imagery data from UAV flight for banana (all varieties) crop class. [2]

2. **Methodology**

After aggregating data, I will identify varieties of banana in different provinces of Rwanda. This will help identifying the area with high and low in number of banana variety.

3. **Analysis**

To start the analysis, I will have to convert the *shapefile* into *Dataframes*, using a python library called ***geopandas***. Since data will contain banana crop varieties, it is a good idea to determine

how many variety of banana crop in dataset, using *data.VARIETY.value_counts ()*. Most of dataset have many columns that might not be relevant for the task. I would like to subset the columns needed for the analysis.

Every GeoDataFrame contains a special "geometry" column. I will next need to understand the geometric objects I will be dealing with either Point, LineString or Polygon.

geometry
POLYGON ((486093.245 4635308.586, 486787.235 4...
POLYGON ((491931.514 4637416.256, 491305.424 4...
POLYGON ((486000.287 4635834.453, 485007.550 4...
POLYGON ((541716.775 4675243.268, 541217.579 4...
POLYGON ((583896.043 4909643.187, 583891.200 4...

Figure 1: Geometry Column

The last part will be to make map or plot for the data. I will creating 3 GeoDataFrames:

a. Banana crops location with Point object,
b. Road or Trail with LineString object *,
c. Provinces and/or Districts boundaries as Polygon object.

* The reason of including road/Trail, is to find how the products will be accessible from both parties either banana farmers and their clients

Proceeding to plotting, I will have to define a base map as provinces/district's boundaries and add other maps to it. In this stage, I can use the normal **plot()** methods or use **interactive maps** with *folium*, a powerful data visualization library in python. [3]. The type of map that will show how density of banana crops varies across districts will be the Choropleth maps.

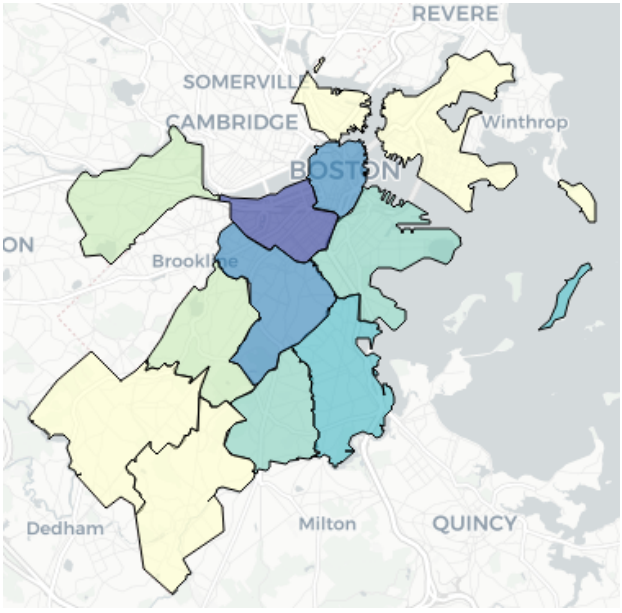

Figure 2: Choropleth map

From maps like the above one, I can present to stakeholders where there is banana crops and provide recommendations of where we can place other farms of banana, why there is presence of banana on some areas other than other places and what to be done to maintain the already built farms, etc.

References

[1] T. Contributor. [Online]. Available: https://whatis.techtarget.com/definition/geospatial-analysis.

[2] J. P. ,. O. ,. U. ,. R. ,. H. B. ,. L. D. S. T. Meghan Hegarty-Craver, "Remote Crop Mapping at Scale: Using Satellite Imagery and UAV-Acquired Data as Ground Truth," *Remote Sensing,* vol. 12, no. Remote Sensing for Agriculture in the Developing World, 2020.

[3] GeeksforGeeks, 03 July 2020. [Online]. Available: https://www.geeksforgeeks.org/stamen-toner-stamen-terrain-and-mapbox-bright-maps-in-python-folium/.