Site Selection Prediction for Supermarkets Based On Multi-Source Spatial Data Using Machine Learning Techniques

by

Kamunya Rebecca Njoki

Project research report submitted to the department of Geomatic Engineering and Geospatial Information Systems for the award of degree of Bachelor of Science in Geospatial Information Science (GIS), 2024.

Department of Geomatic Engineering and Geospatial Information Systems (GEGIS)

DECLARATION

I declare that this project is my own work and has not been submitted by anybody else in any other university for the award of any degree to the best of my knowledge.

Sign…… Date……...

Kamunya Rebecca Njoki

ENC222-0151/2020

Department of Geomatic Engineering and Geospatial Information Systems (GEGIS)

Jomo Kenyatta University of Agriculture and Technology

CERTIFICATION

This project has been submitted for examination with my approval as the candidate's supervisor.

Sign……. Date ……

Dr. Charles Gaya, PhD.

Lecturer, GEGIS

Department of Geomatic Engineering and Geospatial Information Systems (GEGIS)

I

I

Acknowledgements

I would like to express my sincere gratitude to God, my project supervisor Doctor Charles Gaya and the lab technicians for their valuable guidance in making this project a success.

I would also like to extend my gratitude to my family for their support and care throughout my research work.

Abstract

This study examined the spatial distribution of supermarkets in the main urban area of Nairobi whilst selecting the main influencing factors based on the multi-source space data. Subsequently, three regression models were compared, and the best site selection model was found. A comparison was performed between the prediction model functioning with a buffer and without one, and the accuracy of the location model was verified using the MAE, MSE and RMSE validation metrics. The following conclusions were obtained: (1) Supermarkets in the main urban area of Nairobi exhibit a statistically significant clustered distribution in an area within 12 km of the Nairobi CBD, and also around the commercial consumption cluster areas; (2) The Random Forest (RF) model was the best model in this study before and after establishing the buffer zone as it outperformed the GD method and the OLS model in accuracy; and (3) Among the top recommended commercial consumption clusters were The Address, Safaricom House 2, Lion Place, Professor Nelson Awori Centre and Golden Ivy. This study provides crucial insight for supermarket prediction model selection and potential store location selection, which is significant to promoting economic development.

Contents

List of figures

List of tables

Acronyms and abbreviations

1.	RF	Random Forest
2.	GD	Gradient Descent
3.	OLS	Ordinary Least squares
4.	WOK	who owns kenya
5.	WDG	Wadley-Donovan-Gutshaw Consulting
6.	KDE	Kernel Density Estimation
7.	ANNI	Average Nearest Neighbour Index
8.	R-K	Ripley – K Function Curve

1. Introduction

1.1 Background

Kenya has the second most developed retail market in sub-Saharan Africa with about 30 per cent of retail shopping being done in formal outlets (Business Daily, 2012).

The high level of competition experienced in Kenya's retail sector has been a poised difficulty to strategic management of firms. These competitive forces are notably influenced by the severe quest of firms towards achieving improved turnover in sales (Lechner & Gudmundsson, 2014).

Since the retail sector remains a highly competitive and resilient market, opening a supermarket in an area that encourages business presence is essential! (WDG Consulting, 2024)

1.2 Motivation and problem statement

Under this huge competitive pressure, choosing a favorable geographical location to open a supermarket is a necessary task in order to achieve substantial profits**.**

Most selection processes of model-influencing factors depends on the author's subjective ideas and clearly fitting and explaining the influence of these factors on site selection objectives is rather challenging.

Many studies on the locations of commercial stores focus on a single influencing factor or a single forecasting method(Jiaqi Zhao, 2023); therefore, comparing multiple location methods when considering multiple influencing factors is meaningful as it gives a better predictive performance relative to single forecasts.(Ugoh, 2023)

1.3 Justification

The retail distribution sector is facing a difficult time as the current landscape is characterized by ever-increasing competition as seen by Quickmart, Naivas and Carrefour, in the race to expand their footprints in the country, targeting more countries in their aggressive expansion drive.(The Star, 2022)

Retailers are also strategic when it comes to scouting for location for its stores.(WOK, 2023)

An appropriate location strategy thus plays an important role in understanding the success of the business!(Roig N., 2013)

1.4 Research identification and Objectives

The main objective of this research is to select sites for supermarkets based on Multi-source Space data using Machine Learning Techniques. This is achievable through the following specific objectives:

1.4.1 Research objectives

1 To analyze the spatial distribution characteristics of supermarkets.

2 To compare and select best regression methods indicating key factors influencing the location of supermarkets.

3 To propose new sites for supermarkets using machine learning techniques.

1.4.2 Research questions

The following questions are formulated with respect to aforementioned objectives:

- What kind of distribution characteristics do Supermarkets in Nairobi County exhibit?
- What is the best regression method that can predict density of supermarkets accurately?
- What are the recommended sites for opening a supermarket?

1.5 Study outline

This research study is divided into 6 main chapters whereby the first chapter introduces the study by detailing into the background, motivation and problem statement, justifying the problem, objectives and research questions; Chapter 2 contains the reviewed literature which relates to this thesis. Further, Chapter 3 shows the data and methods used in the study with Chapter 4 highlighting the results for the findings from the methods. Chapter 5 discussed on the findings, Chapter 6 concludes and recommends for future research that might not be addressed at this level of geoscientific exploration and expertise.

2. Literature review

2.1 Brief Introduction with Real-World Examples

Research on services, especially in retail, accommodation, and food services, involves location variables with multiple objectives. Determining the optimal location of a service facility is often a complicated task.(Dyah W., 2020)

Spatial location, for many years has been an important factor in fields such as urban planning, business and transportation. Site selection is a critical aspect of strategic planning for a broad spectrum of public and private organizations. Strategic planners are often challenged by difficult spatial resource allocation decisions; thus, determining the best locations for new facilities is an important strategic challenge. Therefore, it is important to identify the challenges and failures that have occurred previously to fully understand and avoid this problem.(Omar A., 2018)

A real-world example can be seen with Shoprite Holdings Ltd, Africa's largest fast-moving consumer goods retailer. In April 2020, the retailer brought its aggressive expansion into Kenya to a sudden halt and announced the closure of its Waterfront branch in Nairobi's Karen Estate, rendering more than 100 workers redundant. This move was part of a wider plan to exit markets that have proved to be unprofitable on the continent(The East African, 2020). Besides economic issues, location is a key factor in why specific branches experience an extreme decrease in sales ultimately causing them to close. Therefore, understanding how to optimally choose new locations with sustainable, increasing sales is a businesses target (Thau, 2014).

Another instance in which the importance of spatial location in the economic space is shown, is with the big retailer, Tuskys. BrandSearch explained the history of Tuskys and their long-standing success, only for Tuskys to announce in February 2020, a restructuring programme that put an unspecified number of employees on the

chopping board, citing a difficult operating environment and poor performance of the business over the past two years characterised by declining sales and customer numbers. (The East African, 2020) Despite the previous successes of Tuskys, the forced closures, as seen in April 2020, illustrate the importance of businesses allocating their facilities in optimum locations that are sustainable in revenue.

Although economic reasons seem to easily explain these closures, as detailed above, they fail to address why specific locations are vulnerable, but not others. Why is this? Simply stated, location plays a crucial role and has a direct effect on the profit and performance; for this reason, it is important to locate the best locations that are sustainable and have high revenue.(Omar A., 2018)

2.2 Early Models for Decision Making in Retail Site Locations

The characteristics of a candidate location may be understood from spatial and aspatial data. When discussing spatial data, the primary concern is the interaction between data pairs. The first law of geography from Tobler states that "Everything is related to everything else, but closer things are more related than distant things". This law premises that when determining a location, service providers must consider the characteristics of the location in their strategies, including the interaction of the location with other locations for example schools, bus stops... as well as the requirements of consumers or other stakeholders.(Dyah W., 2020)

Location determination methods have continued to evolve. In the late 1980s, the use of spatial data grew with the development of Geographical Information Systems (GIS). Some of these early location determination models are the Huff Model and the Gravity Model. The Huff Model is a spatial interaction model that calculates gravity-based probabilities of consumers at each origin location each store in the store dataset. Sales potential can be calculated for each origin location based on knowing and using the population size, their income level, and/or other variables. Hence, the Huff Model depends heavily on the calculation of distance similar to the gravity model (Okunuki & Okabe, 2002). The Gravity Model is used to estimate the amount of interactions between two cities. It is based on Newton's universal law of gravitation, which measures the attraction of two objects based

off their mass and distance (Rodriguez, Jean, 2013) which may not fully conform to the consumer's shopping patterns(Zhao, 2023); these are two of the oldest all-encompassing spatial location models that analyze ideal locations.

Current retailers now have an extensive choice of methods for locational planning. With the emergence of the GIS, the mentioned models have been developed and enhanced with the spatial representation of geo-demographic and retail data based on digitalized cartography, making the spatial interaction models a practical forecasting reality. (Birkin, Clarke & Clarke, 2002). However, a single machine learning method cannot explain the influence mechanism of the influencing factors(Zhao, 2023). This clearly illustrates that spatial planners are acutely aware of the need for multiple computer based methods to help predict successful locations!(Omar A., 2018)

2.3 Spatial Analytics

Spatial co-location pattern mining is a spatial data mining technique for location-based data analysis. This technique represents the relationship between two spatial features that are often adjacent to each other(Kim SK., 2014). This has been applied to various types of spatial data(points, lines and polygons).

The regional co-location pattern mining approach can also be applied to data points by applying Kernel Density Estimation (KDE) surface techniques to each feature; creating a density of features in a neighbourhood around those features. Other various methods of co-location pattern mining have their benefits and inefficiencies in achieving the various objectives designed by the researchers.(Dyah W., 2020)

The sharing of open-source data such as points of interest (POI), OpenStreetMap (OSM), provides massive and accessible data sources for site selection models. For example, Geng Lin et al. analyzed the locations of various retail stores and street centers using POI and street network data.(Zhao, 2023)

2.4 Accuracy Evaluation

The accuracy evaluation and success evaluation of site selection model construction have also attracted the research attention of many scholars because correct site selection may bring very good benefits, while incorrect site selection may bring serious business risks. In the machine learning prediction models of Yuxue Wang and Hui-Jia Yee, MAE , RMSE, MSE and other indicators to measure the performance of machine learning algorithms are mostly used to evaluate the models.

According to BMC Blog(2024), the lower the MSE, the better the Model and 0 means that the model is perfect. RMSE values mean that the model can relatively predict the data accurately and according to Saeedia, a PHD Candidate, RMSE values between 0.2 and 0.5 are well in place. The MSE is sensitive to large deviations of data while the MAE is more robust is outliers and shows the consistency of errors than their sizes according to a source in the internet, thus crucial validation metrics.

2.5 Conclusion

This study will use multiple machine learning algorithms namely Random Forest, Gradient Descent and Ordinary Least Squares utilizing data from multiple sources. The outputs of these machine learning algorithms will then be compared and validated to select the best model that can accurately predict or suggest suitable locations.

3. Materials and methods

3.1 Study area

Nairobi is situated in the south-central part of Kenya, in the highlands at an elevation of about 1,680 meters. It has an estimated population of 4,750,056 and covers an area of 6,784km^2. It continues to feature as one of the leading retail hubs across Sub Saharan Africa with a retail density estimated at 0.14. This figure is higher than most cities in Africa such as Lagos and Accra who have a recorded density of 0.018 and 0.06 respectively. (Estate Intel, 2024)

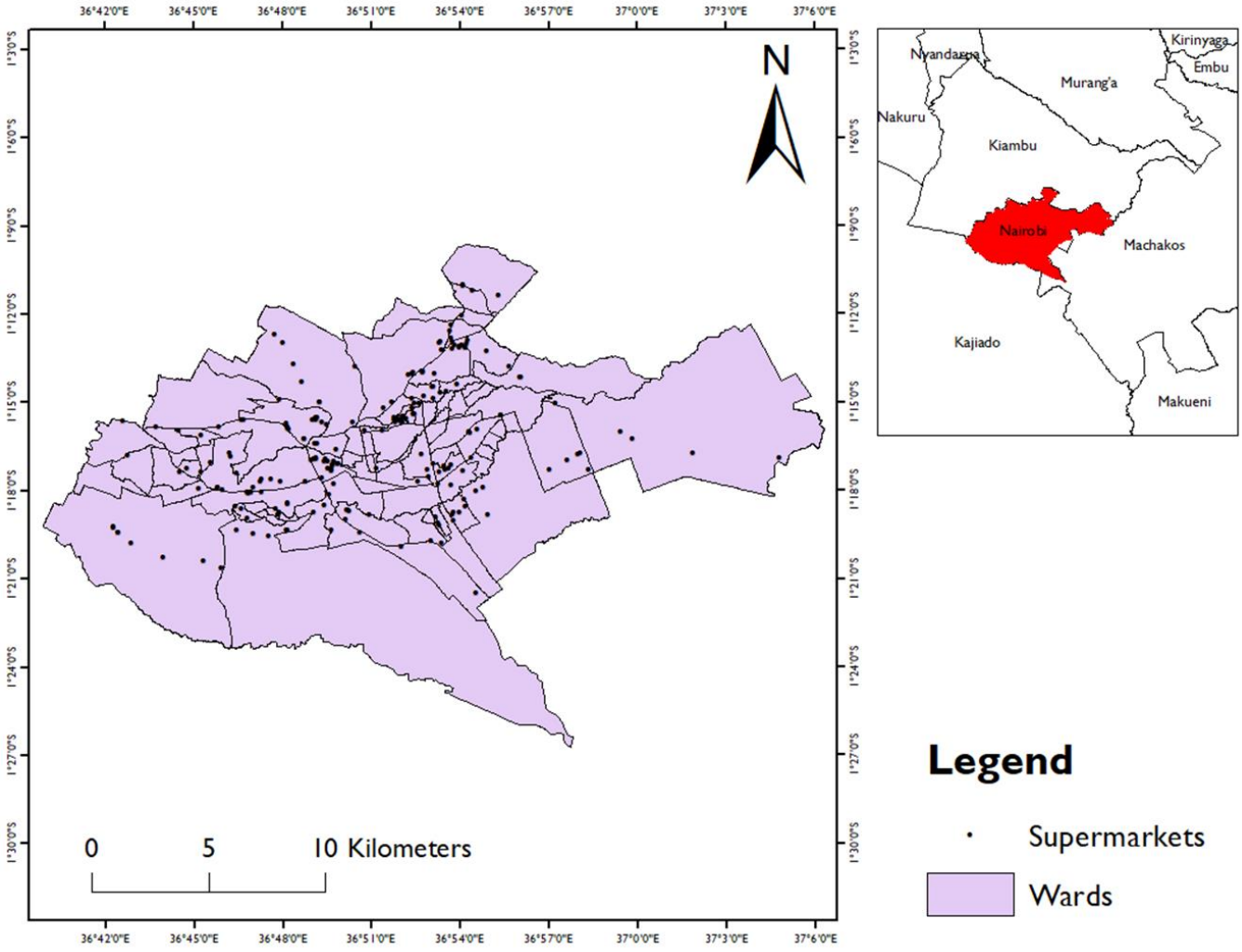


Figure 1 showing study area.

3.2 Data

Four (4) datasets were exploited in this study as displayed in table 1 below.

ITEM	DATA SOURCES	DATA DESCRIPTION	DATA TYPE
POI Data	OpenStreetMap(OSM) (https://www.openstreetmap.org/)	20 Categories of POI data in Main Urban Area	Point
Population Density Data	Humanitarian Data Exchange(HDE) (https://data.humdata.org/)	1km by 1km Spatial Resolution	Raster
Road Network Data	OpenStreetMap(OSM) (https://www.openstreetmap.org/)	Five categories—primary, secondary, tertiary, motorway, and trunk—were identified as major urban roads, and the correlation between store distribution and roads was studied.	Line
Administrative Division Data	Humanitarian Data Exchange(HDE) (https://data.humdata.org/)	Latest version of Nairobi County and its wards.	Polygon

Table 1 showing Data Sources

The following (3) tools were utilized to generate results:

Tool/Material	Role	Availability
R Studio	To perform algorithmic and statistical functions	Freely Available
Excel	To perform statistical functions and generate charts.	Freely Available
Arcmap	To generate maps and perform spatial functions	Freely Available

Table 2 showing Tools and Materials used in the study

3.3 Methodology

3.3.1 Data Preprocessing

Clipping

Clipping a piece of Kenya using the Nairobi county shapefile as the cookie cutter.

Spatial Join

Counting the number of features inside a feature, for example, counting the number of supermarkets in a ward.

Sampling

Randomly selecting a number of sampling units. This study established sampling points on the geometric centre of a ward that was in Nairobi County. The values of characteristic factors were then extracted from the sampling points.

Standardizing

Ensuring that data has a normal distribution and that they are in the same scale or fixed range making it easier for the machine learning algorithms to understand the data.

Selection

So long as the building had the attribute in the field "class" as commercial, it was deemed fit to be a candidate location.

All supermarkets in Nairobi were regarded as homogenous geospatial units, without considering the influence of service quality, operating cost, and profit on location selection.

The sizes of all commercial centres as well as their rent prices were also not considered.

3.3.2 Analysis of Spatial Distribution Characteristics of Supermarkets

The spatial distribution of supermarkets was analysed in 4 ways:

3.3.2.1 Ripley K Function

This is a function that determines whether features exhibit a statistically significant clustering or dispersion within a range of distances.

$$L(d) = \sqrt{\frac{A\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N} k(i,j)}{\pi N(N-1)}}$$

Formula 1 showing how the Ripley K Function works

where:

- d represents distance
- A represents the total area of features
- N represents the total number of features
- K(*i,j*) represents the weight

3.3.2.2 Average Nearest Neighbour Index

This index measures the average distance between a feature's centroid and its nearest neighbour centroid and averages all the nearest neighbour distances

If the average distance is less than that of a hypothetical random distribution, the features are clustered and if the average distance is higher than that of a hypothetical random distribution, the features are dispersed.

$$ANNI = \frac{D_O}{D_E}$$

where:

- D_O is the observed mean distance
- D_E is the expected mean distance

3.3.2.3 KDE and Buffer Analysis

Kernel Density Estimation is a non-parametric method that calculates the density of features in a neighbourhood around those features. It is an important statistical analysis method for extracting the distribution characteristics as well as represent the agglomeration area of dotted elements in space. If the location is closer to the core element, the kernel density will be greater, indicating a higher aggregation degree of the midpoint in this region. (Zhao, 2023)

$$f(x) = \sum_{i=1}^{n} \frac{1}{r^2} k\left(\frac{x - x_i}{r}\right)$$

Formula 2 showing how KDE works

where:

- f(x) is the kernel density at the point *x*
- n represents the number of elements in the range of distance *r* from the point *x*
- k is the spatial weight coefficient
- r is the distance decay threshold

In buffer analysis, multiple buffer rings are created at specified distances around features. Buffer zones help to identify the influence of a spatial object on its surrounding features(Zhao, 2023). For a given object, its buffer(*P*) can be defined as:

$$P = \{x \mid\mid d(x, A) \leq r\}$$

Formula 3 *showing how a buffer zone is defined*

where:

- d is the distance between point *x* and point *A* in space
- r is the neighbourhood radius

There are three common patterns in buffer analysis: the point, line, and surface. By establishing a buffer zone, the influence ranges of geographical elements can be expressed. (Ling, 2023). In this paper, a multilevel buffer zone was centered on Nairobi CBD to study the spatial distribution characteristics of supermarkets and took the constraints of main roads in the city as a buffer zone to screen out some areas for better site selection. This was applied

in creating multiple buffers with specific distances from the Nairobi CBD and along the main roads to screen out locations that were outside the 150m buffer.

3.3.3 Best Regression Model selection

3.3.3.1 Construction of influencing factors

Based on a study by Tuzla, Hayri, and Sharon Cobb, I preliminarily constructed 20 factors (Figure 2) that may affect the location of supermarkets in 6 ways; the explanations of each factor are shown in Table 3

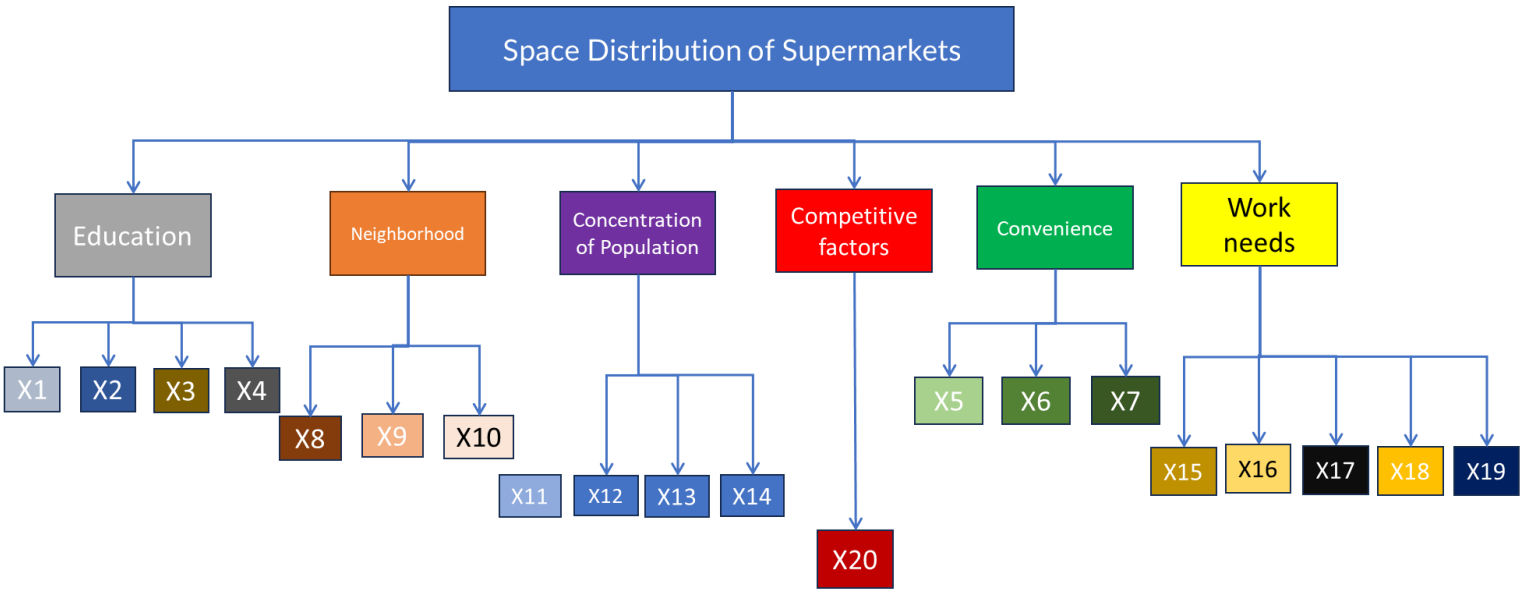


Figure 2 showing 20 indicators that could affect the site of supermarkets

Feature	Name	Category
X1	Colleges	EDUCATION
X2	Kindergarten	
X3	Schools	
X4	Universities	
X5	Bus stops	CONVENIENCE
X6	Bus Stations	
X7	Taxi	
X8	Hospitals	NEIGHBORHOOD
X9	Hotels	
X10	Gas Stations	
X11	Parks	CONCENTRATION OF POPULATION
X12	Sport Centers	
X13	Community Centres	
X14	Population per sq km	
X15	Bank	WORK NEEDS
X16	Police	
X17	Post-Office	
X18	Office Buildings	
X19	Commercial centres	
X20	Supermarkets	COMPETITIVENESS

3a

Statistics of the 6 categories of POI	
Education	1731
Neighborhood	579
Concentration of Population	213
Competition	213
Convenience	328
Work needs	802

3b

Table 3a showing 20 indicators mentioned in figure 4 and their stats in Table 3b.

Binary classification was performed to explore the feasibility of the location selection method and determine whether there was a supermarket in a ward. The suitability of the above indicators was then evaluated, and factors with a small influence on site selection were eliminated through a process called *pruning*.

Road data covered was considered separately as factors independently of the 20 influencing factors

3.3.3.2 Construction of road buffer zone

In cities, most of the store facilities are distributed along the streets (Han, 2019) and human activities are also constrained by the roads, while the traditional analysis method is based on the whole region, which has some limitations. For example, the traditional analysis method generally adopts the globally consistent parameter model for analysis but ignores the spatial non-stationarity of the influencing factors when applied to the analysis of commercial facility vitality (Teng W., 2017). Therefore, this study considered the comparison of multivariate

regression models taking into account road constraints. Based on OSM data, multilevel buffer zones for key roads in the main urban area of Nairobi were established by setting appropriate distance intervals through analysis.

Roads are linear elements; therefore, the construction of a multilevel road buffer can filter out a part of the area, reduce the bias generated in the construction model, and may have the potential to improve the accuracy of the location model (Han, 2019). The distribution of supermarkets around a road by constructing a multilevel road buffer and study the influence of road buffer construction on the accuracy of the prediction model was analysed.

3.3.3.3 Random Forest

This is a famous machine learning algorithm that uses supervised learning methods. RF models have high accuracy among many machine learning models used for location problems according to research done by Hui-Jia Yee et al. Thus this method was chosen to be compared with other two regression methods.

Training samples are randomly selected to establish multiple decision trees, which constitute a base evaluation sequence {*h1(X), h2(X), h3(X), . . . , hk(X)*}, and the order or average of the prediction results of each base evaluation sequence determines the prediction results of the unknown sample.

$$H(x) = argmax_y \sum_{i=1}^{k} I(h_i(X) = Y)$$

Formula 4 *showing how RF works*

where:

- H(x) represents the final result of the model
- I is the schematic function
- h_i is the base evaluator
- Y is the target variable

When *Y* is a classification variable, the model is used to solve the classification problem. When *Y* is a continuous variable, the model is used to solve the regression problem.(Li, 2013) This basically judges the importance of features by comparing the prediction accuracy of the model before and after adding noise. If the model prediction accuracy is greatly reduced, the feature is important. RF cannot obtain certain equations as traditional models do, and it usually uses the data to evaluate the model. Usually, the established model is applied to the

training and test data, and the model evaluation results are obtained. 60% of the data was used for training and the remaining 40% for testing. In RF classification problems, the error is the error rate of the classification, and the model performance is often evaluated based on a confusion matrix.(Zhao, 2023) Recall, Precision, and F1 were selected in this study to evaluate the model performance with the following formulae respectively:

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{7}$$

Where:

- TP is the number of *true-positive* classes
- FP is the number of *false-positive* classes
- FN is the number of *false-negative* classes.

3.3.3.4 Gradient Descent

This is an optimization algorithm that seeks to minimize cost; solves the cost function so that the cost function loses the minimum value. Since it can go through several iterations, it is widely used in solving optimization problem for model parameters. This study will use GD to optimize the OLS parameters, and compare the accuracy of the results with the traditional OLS regression method.

Suppose that the linear regression function is:

$$h_\theta(x) = \sum_{i=0}^{n} \theta_i x_i = \theta^T x \tag{8}$$

and $x_0 = 1$,

$$J(\theta) = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2 \quad (9)$$

Where:

- x is a feature
- θ is the regression coefficient of this feature
- n is the number of features
- m is the number of samples
- h_θ (x) is the predicted value
- y (i) is the actual value
- J(θ) is the cost function

For the regression coefficient *θ* of j, the loss function gradient is:

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m}\left[\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)x_j^{(i)}\right] \quad (10)$$

To minimize the cost function loss value, iteration can be performed by:

$$\theta_j := \theta_j - \alpha\frac{\partial}{\partial\theta_j}J(\theta) \quad (11)$$

Where:

α is the *learning rate* used to control the *step size* of the GD

As the number of iterations increases, the parameter reaches a stable point and no longer changes, yielding the determined regression coefficient.(Zhao,2023)

3.3.3.5 Ordinary Least Squares

This simply creates a single regression equation that the process takes. It minimizes the sum of squares of the difference between the predicted value and the actual value by solving a set of unknowns. As a classical and stable linear regression model, this model was used as a benchmark to compare with the two other machine learning regression models.

Suppose that the linear regression function is:

$$h_\theta(x) = X\theta \tag{12}$$

Where:

- hθ (x) is the *m* × *1* dimension vector
- θ is the *n* × *1* dimension vector
- X is the *m* × *n* matrix
- m is the number of samples
- n represents the number of features

The loss function of θ is defined as:

$$J(\theta) = \frac{1}{2}(X\,\theta - Y)^T(X\,\theta - Y) \tag{13}$$

where Y is the output vector of the sample, and the dimension is *m* × *1*

According to the principle of least squares, this means that:

$$\frac{\partial}{\partial\theta_j} J(\theta) = 0 \tag{14}$$

Thus after finishing, we find:

$$\theta = \left(X^T X\right)^{-1} X^T Y \tag{15}$$

3.3.4 Precision Evaluation of Model

For all three (3) regression models, the mean squared error (MSE), the root mean squared error (RMSE) and the mean absolute error (MAE) were used to evaluate the models.

The calculation method is as follows:

$$MSE = \frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2 \quad (16)$$

$$RMSE = \sqrt{MSE} \quad (17)$$

$$MAE = \frac{1}{m}\sum_{i=1}^{m}|y_i - \hat{y}_i| \quad (18)$$

Where:

- y_i is the *true* value of the table sample
- $\hat{y}_i$ represents the *predicted* value of the sample

RMSE is the square-root value of the MSE. The MAE is the mean value of the absolute error, which can reflect the actual situation of the predicted error. The smaller the values of the above three indicators, the higher the accuracy of the model.

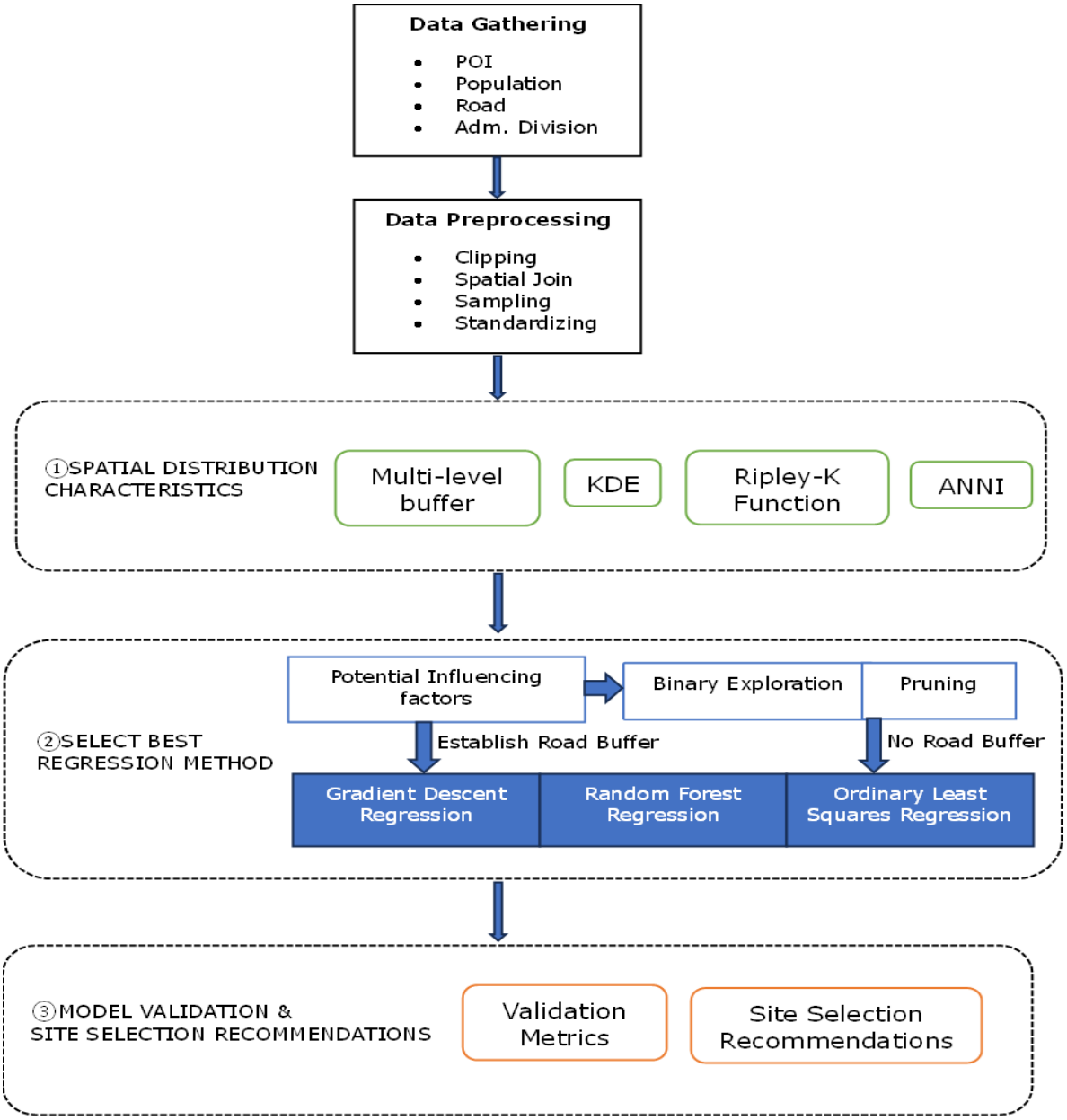


Figure 3 showing the methodology

4. Results

4.1 Spatial Characteristics

4.1.1 Ripley – K Function

Ripley – K function determines whether features exhibit a statistically significant clustering or dispersion within a range of distances

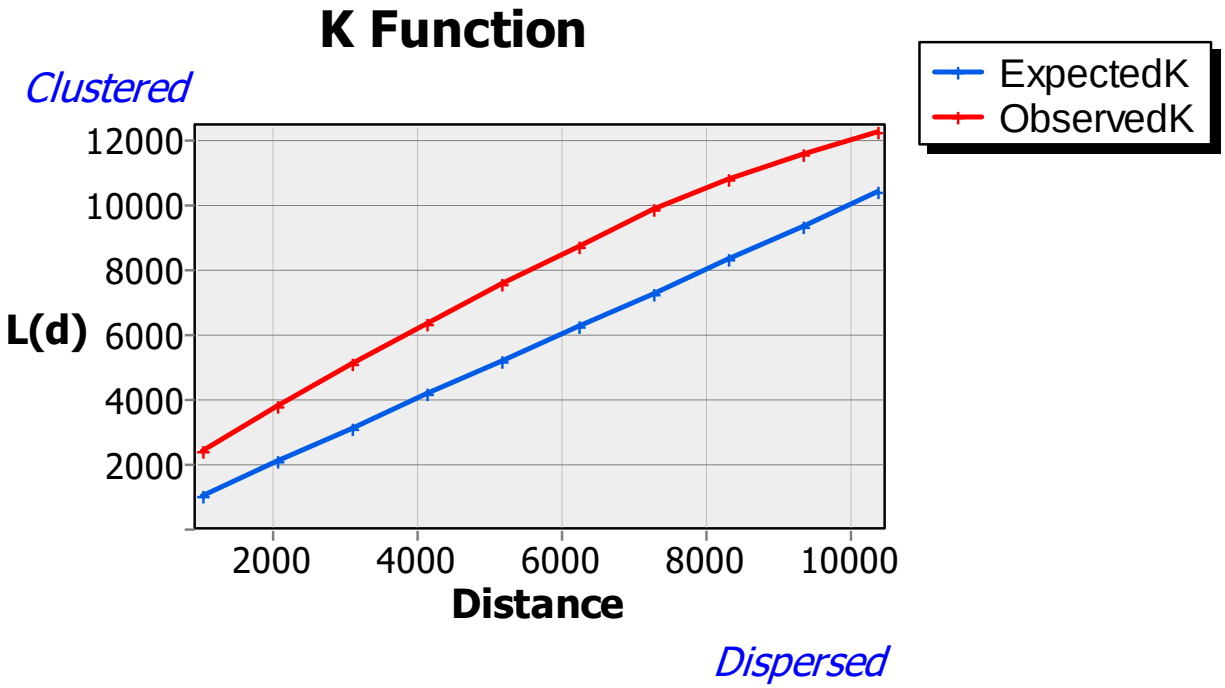


Figure 4 showing Ripley K Function curves of supermarkets

As can be seen from the above figure, the observed values of R-K function curves of all Supermarkets are above the expected value, which indicates that supermarkets show agglomeration distribution in Nairobi, that the distribution is significant, and that the supermarkets exhibit a clustered pattern.

4.1.2 Average Nearest Neighbour Index

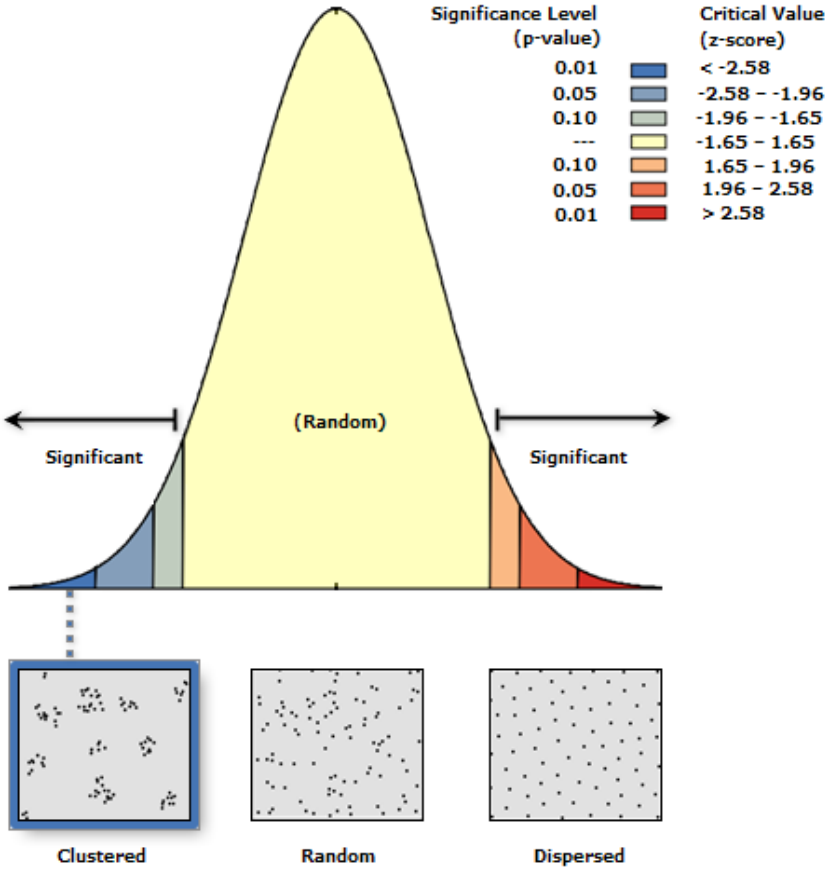


Figure 5. Average Nearest Neighbor Diagram

Project	Supermarkets
Average Observation Distance	521.6166 Meters
Expected observation distance	959.6365 Meters
Nearest neighbor ratio R	0.543556
Z-score	-12.744060
P-score	0.000000

Table 4. Average Nearest Neighbor Index results Diagram

The average distance is less than the expected distance showing that the distribution of features is clustered. Nearest neighbor ratio R value is less than 1 indicating that spatial agglomeration degree is high and exhibits clustering, thus consistent with the R-K Curve.

4.1.3 KDE and Buffer Analysis

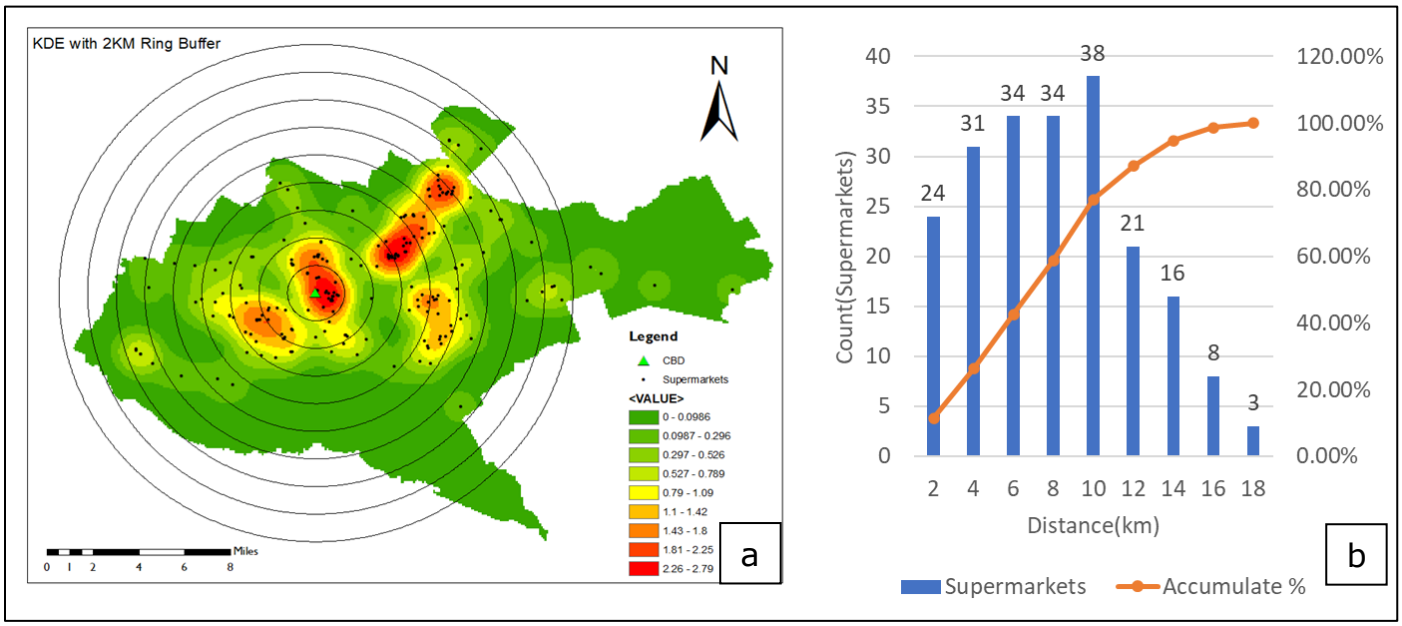


Figure 6. Spatial distribution characteristics of Supermarkets in the study area: (**a**) the result of setting up multiple buffer zones every 2 km outward from the neutral point, which is superimposed on the kernel density analysis results for the Supermarkets; (**b**) the number and cumulative percentage of Supermarkets in each buffer.

For supermarkets in the main urban area of Nairobi, a multilevel buffer zone was established every 2 km in an outward direction, with the Nairobi CBD, being the center.

A statistical analysis of the number of supermarkets in each buffer zone was performed(Figure 6b). From the cumulative number of supermarkets at different distances from the central point (Figure 6b), it was found that more than 86% of the supermarkets in the main urban area of Nairobi are concentrated in an area within 12 km of the central point.

Because a single central buffer cannot represent the true aggregation range(Dolega, 2016) overlaying of the KDE (Figure 6a) results with the multilevel buffer was established.

Through the observation of kernel density interpolation images, it is found that the Supermarkets distribution features had obvious agglomeration effects, and multiple regions with high kernel density values were generated within 2–10 km of the central point. The distribution of high values was consistent with the conclusion of the buffer analysis, the Ripley-k function and Average Nearest Neighbor.

4.2 Best Regression Method Selection

4.2.1 Binary Classification

	Sampling Number	Recall	Precision	F1
Training	60	1	1	1
Testing	25	0.7	0.778	0.737

Table 5. Classification Accuracy

Table 5 shows that the average test accuracy rate was 77.8%, and the recall rate was 70%, both of which reached quite a high level, and the average F1 score was 0.737, which reflected that the precision and recall rate also reached a good level after the average reconciliation. Therefore, it was found that the sampled data achieved good performance when they were used for binary classification of the presence and absence of supermarkets; thus, they could be used for multivariate regression and predictions of supermarket density based on the sampled data.

4.2.2 RF Regression Results

Regression was conducted based on the RF algorithm for the sampled data taking the supermarket KDE as the dependent variable and the other 20 possible influencing factors as independent variables for regression.

The results are shown in Table 6 and Figure 7

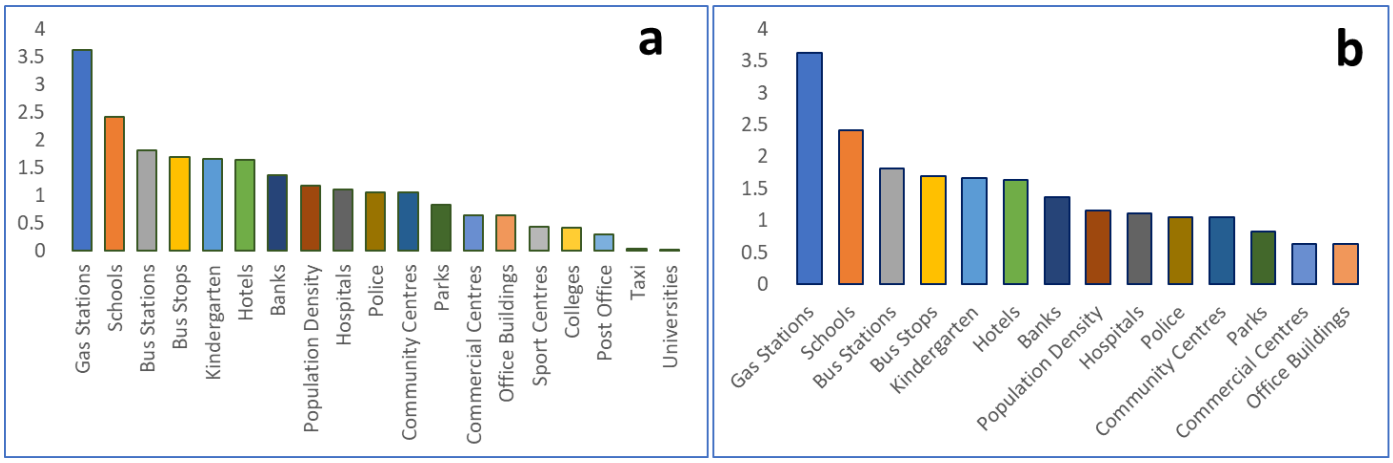


Figure 7. Significant results of RF regression models (not establishing a road buffer): (a) unpruned; (b) pruned;

4.2.3 GD Regression Results

Regression was conducted based on the GD algorithm for the sampled data taking the supermarket KDE as the dependent variable and the other 20 possible influencing factors as independent variables for regression.

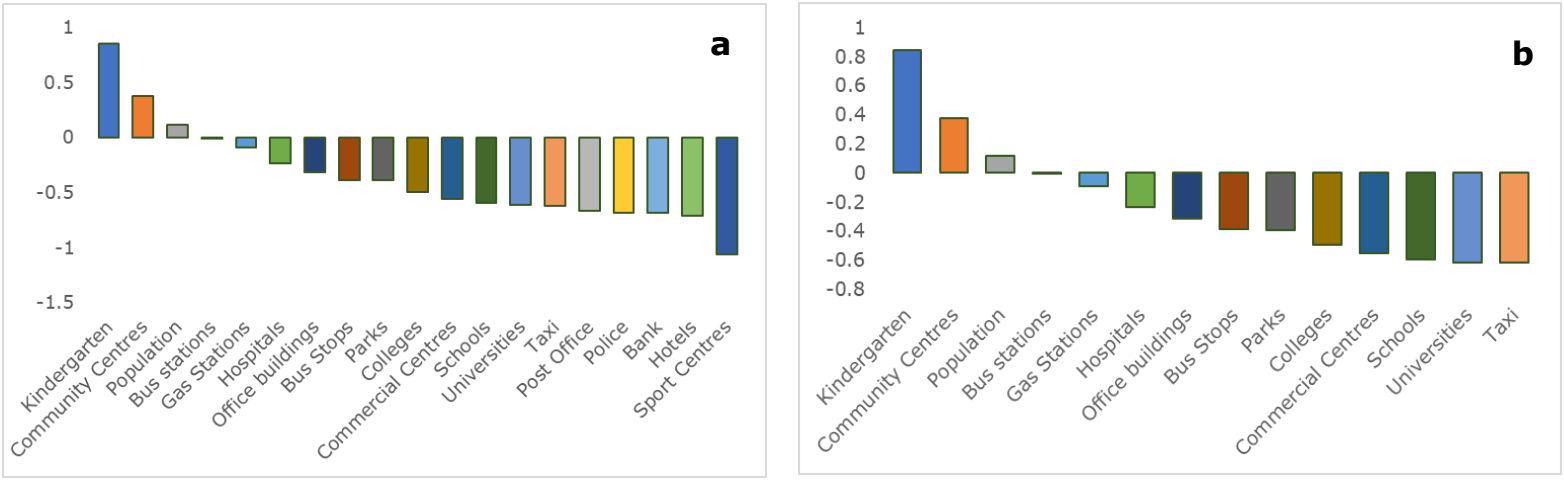


Figure 8. Significant results of GD regression models (not establishing a road buffer): (a) unpruned; (b) pruned;

4.2.4 OLS Regression Results

Regression was conducted based on the OLS algorithm for the sampled data taking the supermarket KDE as the dependent variable and the other 20 possible influencing factors as independent variables for regression.

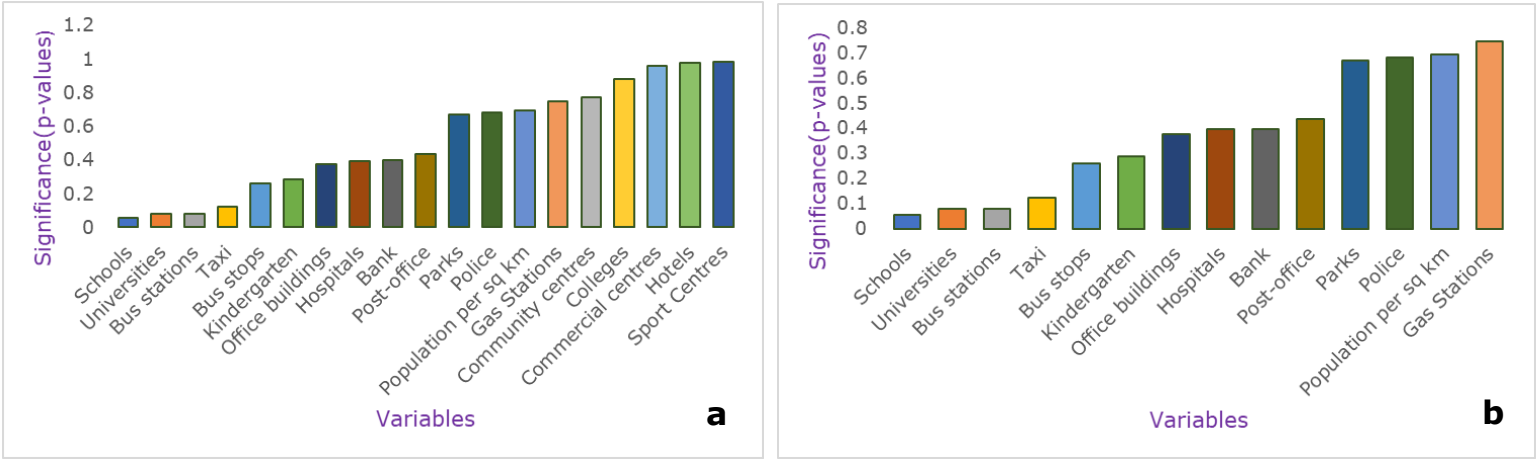


Figure 9. Significant results of OLS regression models (not establishing a road buffer): (a) unpruned; (b) pruned;

4.2.5 Model Accuracy Before and After Pruning

Table 6 shows that the average MSE, RMSE, and MAE for RF were 0.0879, 0.2964, 0.2128 and 0.7251, respectively, i.e., small values, which shows that the expected difference between the predicted value and the actual value was quite small. The validation results showed that the sampled data also achieved high accuracy when used for multivariate regression.

		MSE	RMSE	MAE
Before Pruning	RF	0.0879	0.2964	0.2128
	OLS	0.2708	0.5204	0.3908
	GD	0.4569	0.6760	0.3705

Table 6 showing Model Accuracies before pruning

		MSE	RMSE	MAE
After Pruning	RF	0.0907	0.3012	0.2173
	OLS	0.2785	0.5277	3.8251
	GD	0.4569	0.6760	0.3705

Table 7 showing Model Accuracies after Pruning

Regarding the characteristic importance of influencing factors using RF, Gas Stations(X10), Schools(X3), Bus Stations(X6) were the main factors that had a significant impact on the spatial distribution of supermarkets. Sport centers(X12), Colleges(X1), Post-offices(X17), Universities(X4) and Taxi Stands(X7) were five factors that had low importance. In order to simplify the model and improve the applicability of all 3 models, the five influencing factors whose average feature importance was less than 0.6%, were eliminated.

After eliminating the influencing factors where the average importance of the characteristic was too small, 14 influencing factors were taken as the independent variables, and the supermarket KDE was taken as the dependent variable. RF model regression, linear

regression based on the GD method, and OLS regression were selected to predict the kernel density of the supermarkets.

4.2.6 Comparison of Multivariate Regression Accuracy of the Main Road Network Constraints

The results of constructing a multilevel road buffer are shown in Table 8 and Figure 11. The findings reveal, similar to most commercial facilities, supermarkets were mainly distributed along roads.

The zoning statistics show that about 64% of the supermarkets were distributed in the buffer zone 150 m away from the main road network.

Therefore, sampling points within the buffer zone 150 m from the main road were selected for further regression analysis. Similarly, comparative analysis of the three regression methods was still used after establishing the road buffer zone. The experimental results were obtained, as shown in Table 9.

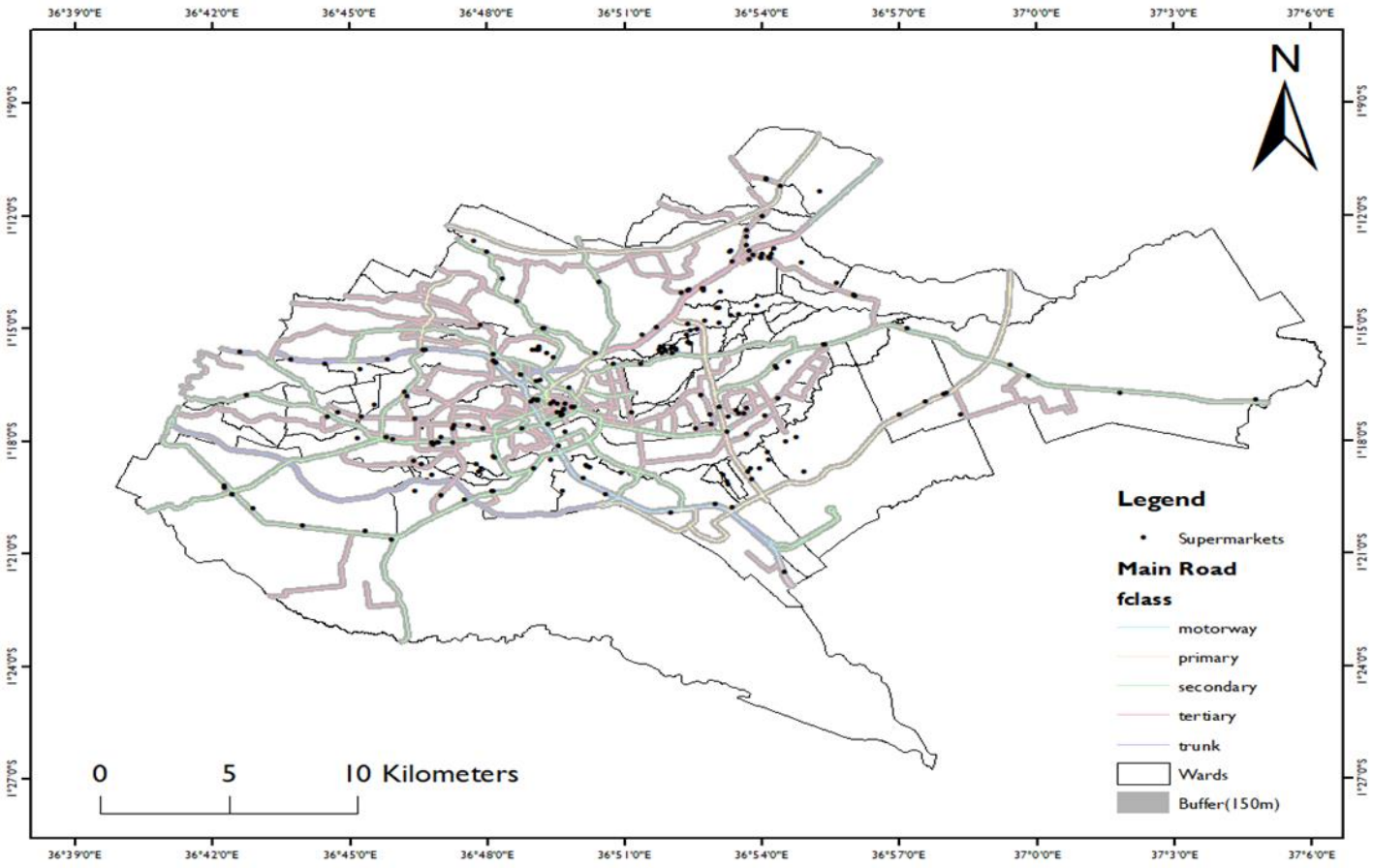


Figure 10 showing the main roads and the 150 m buffer zone

Distance/m	(0,50)	(50,100)	(100,150)	(150,200)	(200,+∞)
Counts	76	36	24	6	71
Cumulative %	36%	53%	64%	67%	100%

Table 8 representing cumulative percentages of supermarkets

		MSE	RMSE	MAE
After Buffer Analysis	RF	0.0891	0.2985	0.2123
	OLS	0.2770	0.5263	3.8315
	GD	0.4569	0.6760	0.3705

Table 9 representing model accuracy after buffer analysis

The results showed that after the regression prediction of buffer processing, the accuracy of the three models changed as follows:

- For machine learning regression based on RF and the OLS method, the accuracy of RMSE and MAE increased.
- For the GD method, nothing changed.
- The regression effect of machine learning based on RF was the best.

Thus far, the three regression methods were compared under the conditions of buffer analysis and non-buffer analysis comprehensive analysis results of the model prediction accuracy were achieved.

The analysis found that the RF method outperformed the GD and OLS methods in the buffer analysis and non-buffer analysis. Between the RF model and the OLS model, the prediction accuracy of RF also outperformed OLS, which indicates that the site selection prediction in this paper may be more biased toward nonlinear influence(Xu, 2022), which is consistent with the discussion of Michael Nwogugu et al.. Within the context of observing the similar distributions of various regression features, the RF model, with better generalization ability and higher prediction accuracy under the two analysis conditions, was chosen as the basic method of my site selection model.

4.3 Site Selection Recommendations

4.3.1 Distribution relationship between the commercial Consumption Agglomeration area and supermarkets

In light of the characteristics of a high consumption rate and high flow of people in commercial consumption gathering areas, commercial centres were selected in the main urban area of Nairobi for site selection prediction and verification in order to provide suggestions for supermarkets locations and the sustainable development of the spatial pattern of commercial consumption.

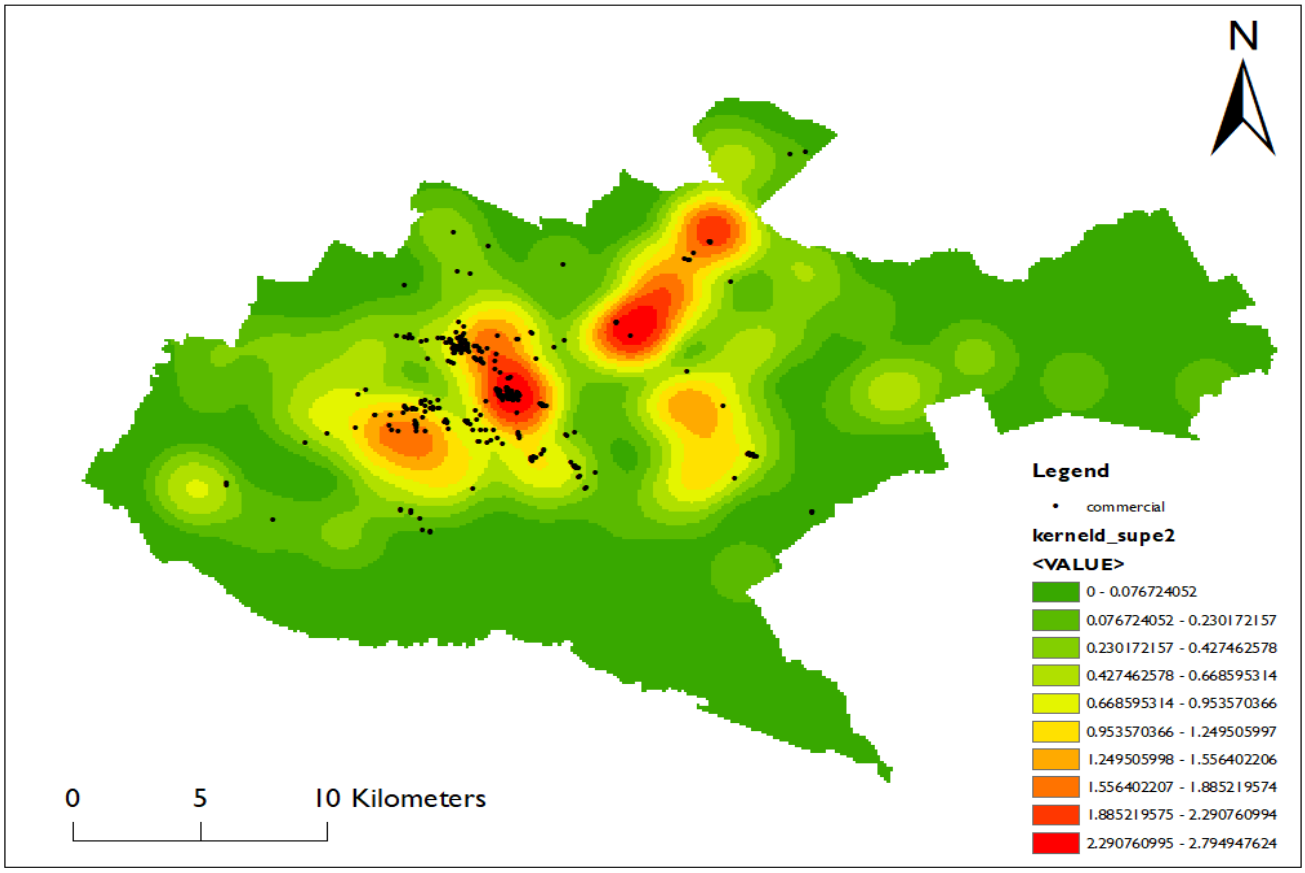


Figure 11 showing commercial consumption agglomeration areas in Nairobi.

The average kernel density of supermarkets in all commercial consumption agglomeration areas was 1.340284094

In addition, there were 120 supermarkets in the buffer zone of the commercial cluster area, accounting for 56.34% of all supermarkets in the main city.

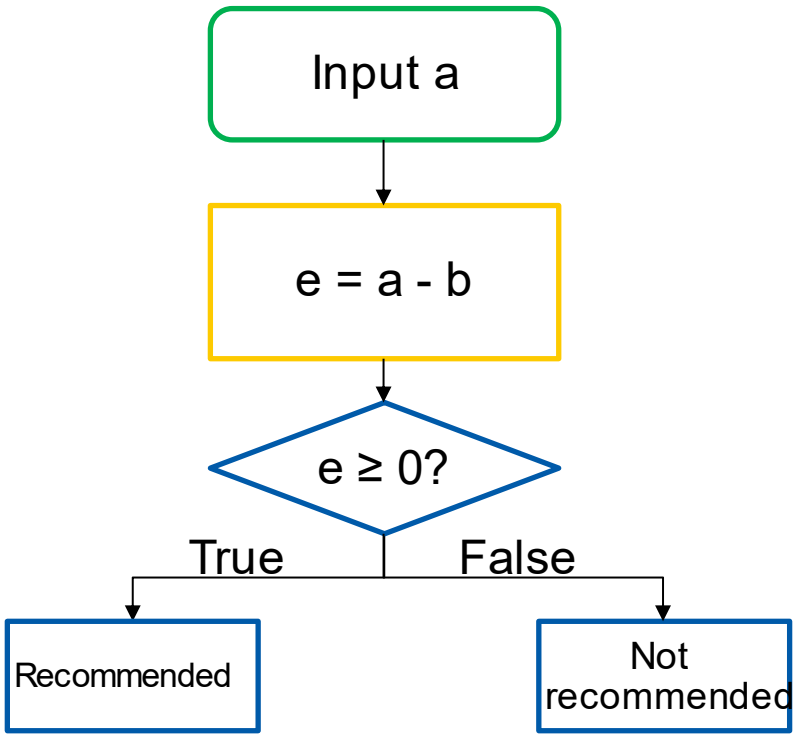
Therefore, it can be considered that the commercial consumption agglomeration area is closely related to the kernel density of Nairobi's supermarkets, as shown by the fact that the commercial consumption agglomeration area highly coincided with the high-value kernel density of supermarkets, and the commercial consumption cluster area had a relatively strong agglomeration effect on supermarkets.

4.3.2 Predicting the success of commercial consumption clusters

Commercial consumption clusters are typical areas suitable for the location of supermarkets and can be used to verify the practicality of the model. (Zhao, 2023)

As shown in Figure 12, by subtracting the predicted kernel density value from the actual kernel density value, it could be determined whether the site selection was recommended or not.

If the difference was greater than 0, site selection was recommended. If the difference was less than or equal to 0, site selection was not recommended.



a : Predicted kernel density of supermarkets

b : True kernel density of supermarkets

Figure 12 showing the method for judging the success of the prediction

4.3.3 Site Selection Advice

The RF model and the POI data were both used to predict the site selection suitability of supermarkets in the commercial consumption agglomeration areas in Nairobi, providing site selection suitability suggestions.

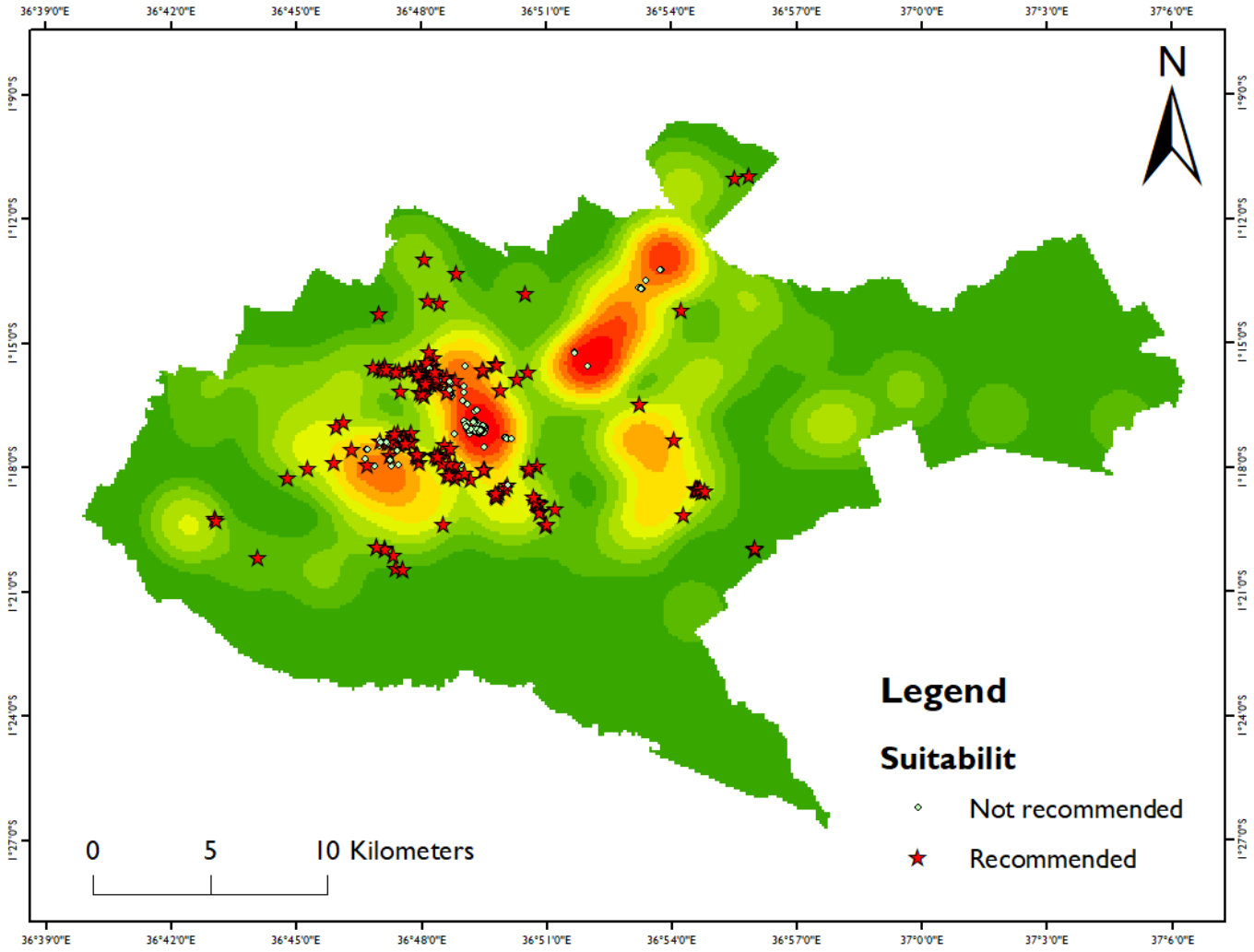


Figure 13 showing Recommended site selection area

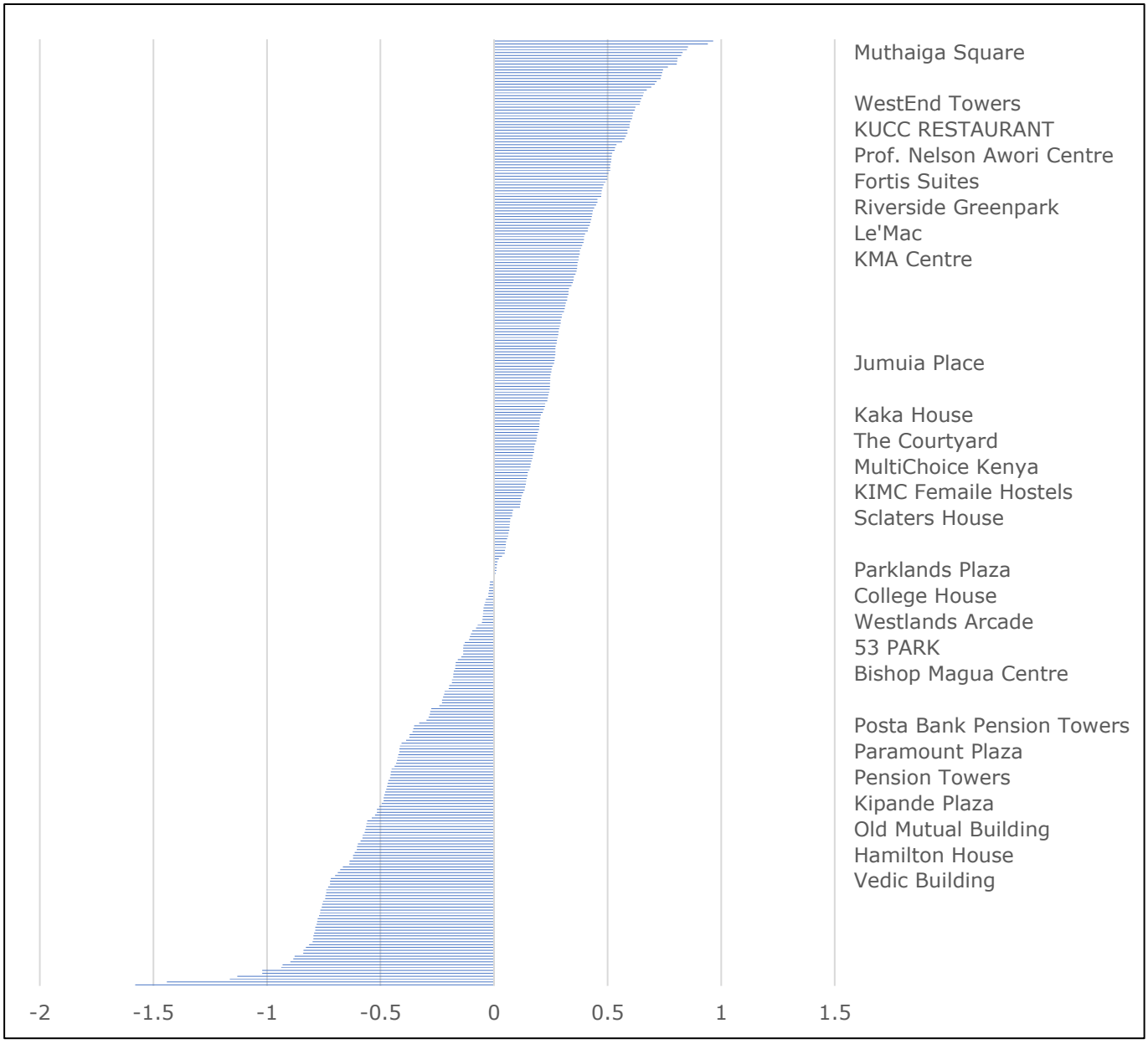


Figure 14 showing location suitability of the commercial centers

The larger the value, the stronger the suitability and the higher the recommendation degree. The recommendation degree of each commercial consumption cluster is shown in Figure 14. The top 10 are Muthaiga Square, Harambee Hall, The Address, Professional Centre, Muthaiga Business Centre, Milestone Business Centre, The Stables Karen, Good Man Tower, The Arches, and WestEnd Towers. Established commercial centers such as Kenya Charity Sweeptake House, Transnational Plaza, and Hamburg House are the least recommended, reflecting the intense pressure on Supermarkets in these areas.

4.3.3.1 Data Validation

This had to be validated by overlaying the current supermarkets on the commercial centers and getting the number of those that were both recommended and had at least a supermarket.

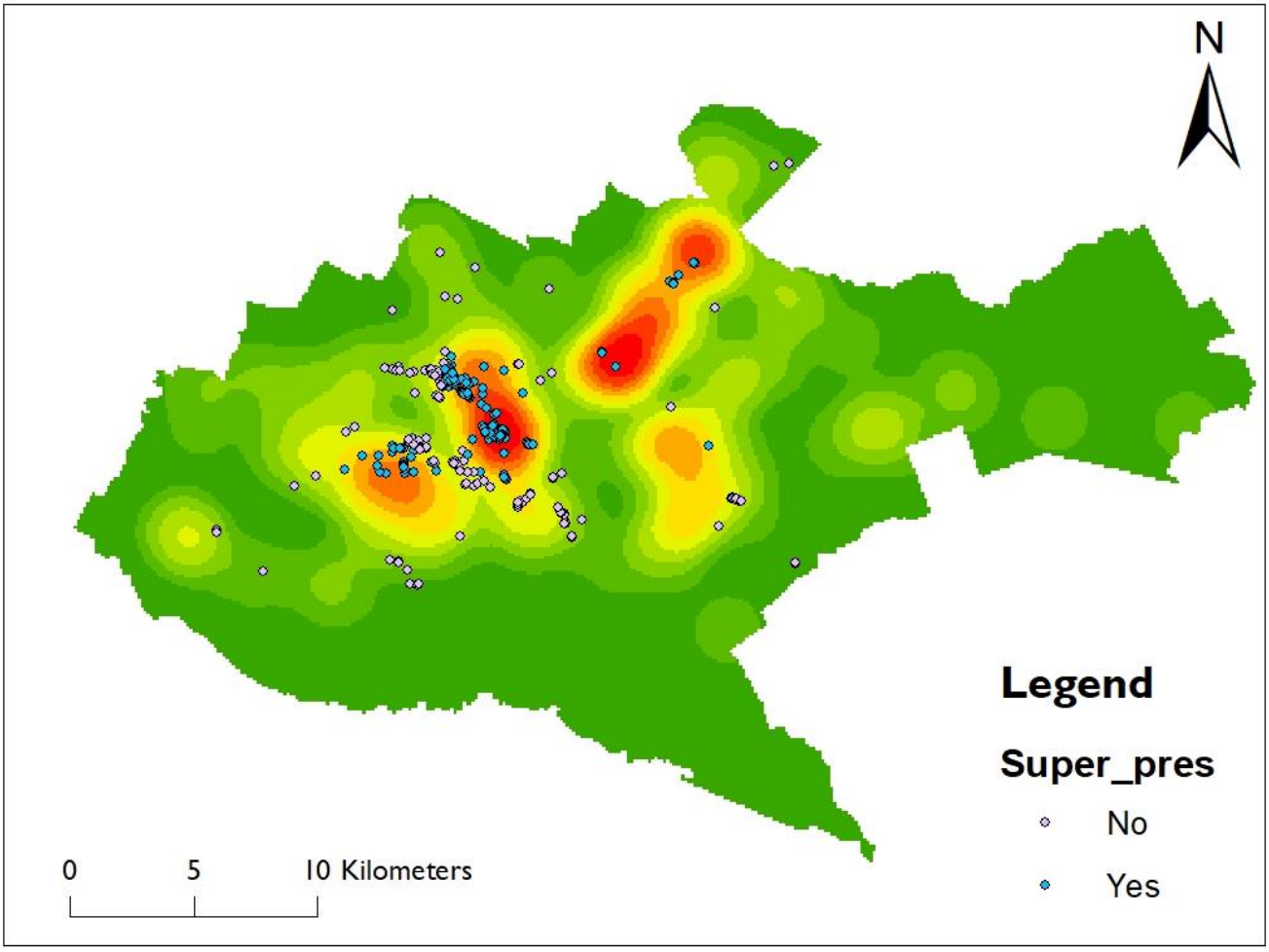


Figure 15 showing commercial centers that had supermarkets and those that did not.

It was found that 139 commercial centers had at least one supermarket while 190 others did not. Those that had supermarkets were further examined by checking whether they were recommended or not.

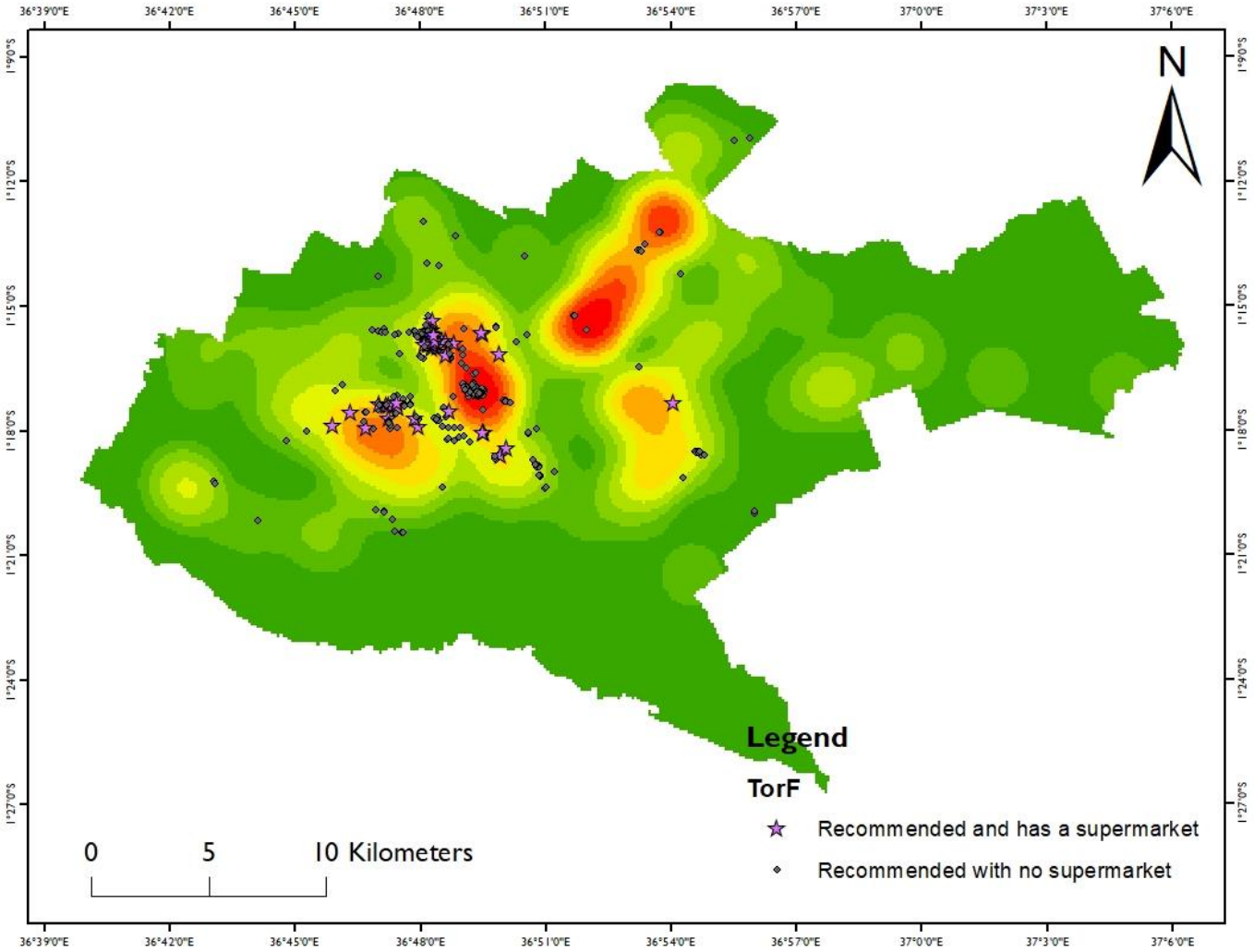


Figure 16 showing recommended sites that had supermarkets and those without.

55 commercial centers, accounting for 16% of all 329 commercial facilities, were recommended and had supermarkets and were located in high kernel density values of supermarkets mostly around the CBD.

5. Discussion

According to BMC Blog(2024), the lower the MSE, the better the Model and 0 means that the model is perfect.

Since my dataset consists of a range of values between 0 and 2(As seen from the kernel density image in figure 6), the RMSE values mean that the model can relatively predict the data accurately according to Saeedia, a PHD Candidate, who states that RMSE Values between 0.2 and 0.5 are well in place.

The MSE is sensitive to large deviations of data while the MAE is more robust to outliers and shows the consistency of errors than their sizes according to a source in the internet, thus essential when evaluating the location methods.

According to the results of the supermarket location analysis, the selected model recommends fewer commercial centers close to the CBD(Ferreira, 2018).

A possible reason is that most commercial consumption centers are in areas with a high kernel density of supermarkets. If businesses set up supermarkets near the Nairobi CBD, they will face great competitive pressure.

6. Conclusions and Recommendations

6.1 Conclusions

- ✓ Supermarkets in the main urban area of Nairobi have obvious spatial distribution characteristics;
    - Present a statistically significant clustered distribution around the commercial consumption cluster areas.
    - Mainly distributed in areas less than 12 km away from the center of the city and no more than 150 m away from the main roads.
    - Present a high correlation with the commercial consumption cluster areas.
- ✓ The RF model was the best model in this study before and after establishing the buffer zone; it outperformed the GD method and the OLS model in accuracy.

- ✓ Among the top recommended commercial consumption clusters were The Address, Safaricom House 2, Lion Place, Professor Nelson Awori Centre and Golden Ivy.

6.2 Recommendations

- ✓ A website showing the various commercial centres indicating their suitability would serve as a valuable tool to supermarket practitioners and researchers.
- ✓ The model can be applied to a specific supermarket to obtain insights.
- ✓ Comprehensive consideration of the other factors such as degree of commercial aggregation, market saturation, land rates and others will improve the prediction accuracy of the location selection model.
- ✓ Furthermore, adding these factors into the site selection model to make it more in line with the actual requirements of site selection is also a problem worth further study in the future.

7. References

Aboulola, O. I. (2018). GIS spatial analysis: A new approach to site selection and decision making for small retail facilities. The Claremont Graduate University.

Business Daily(2012). Survey Ranks Kenya as Africa's Second Largest Retail Market. Business Daily. Accessed on 25 December 2023 at https://www.businessdailyafrica.com/bd/markets/survey-ranks-kenya-as-africa-s-second-largest-retail-market--2006802#google_vignette

Dolega, L., Pavlis, M., & Singleton, A. (2016). Estimating attractiveness, hierarchy and catchment area extents for a national set of retail centre agglomerations. *Journal of Retailing and Consumer Services*, *28*, 78-90.

Ferreira, J., & Ferreira, C. (2018). Challenges and opportunities of new retail horizons in emerging markets: The case of a rising coffee culture in China. *Business Horizons*, *61*(5), 783-796.

Ferreira, J., Ferreira, C., & Bos, E. (2021). Spaces of consumption, connection, and community: Exploring the role of the coffee shop in urban lives. *Geoforum*, *119*, 21-29.

Huang, Q., Yang, B., Xu, X., Hao, H., Liang, L., & Wang, M. (2022). Location selection and prediction of Sexy Tea Store in Changsha city based on multi-source spatial data and random forest model. *J. Geo-Inf. Sci*, *24*, 723-737.

Kim, S. K., Lee, J. H., Ryu, K. H., & Kim, U. (2014). A framework of spatial co-location pattern mining for ubiquitous GIS. *Multimedia tools and applications*, *71*, 199-218.

Lin, G., Chen, X., & Liang, Y. (2018). The location of retail stores and street centrality in Guangzhou, China. *Applied geography*, *100*, 12-20.

Muchere, P. N. (2014). *Marketing Strategies Adopted to Gain a Competitive Advantage by Supermarkets in Kakamega Town, Kenya* (Doctoral dissertation, University Of Nairobi).

Murithi, W., & Kah, S. (2024). Tuskys Supermarkets: the good, the bad and the ugly in the Kago family business. In *Case Studies in Family Business* (pp. 49-62). Edward Elgar Publishing.

Rowe, W. (2018). Mean Square Error & R2 Score Clearly Explained. BMC. Available online: https://www. bmc. com/blogs/mean-squared-error-r2-and-variance-in-regression-analysis (accessed on 15 June 2023).

The East African(2020). Tuskys in the footsteps of collapsed Nakumatt and Uchumi. The East African. Accessed 28 July 2024 at https://www.theeastafrican.co.ke/tea/business/tuskys-in-the-footsteps-of-collapsed-nakumatt-and-uchumi-1936046#google_vignette

The Star(2023). Quickmart continues to revolutionize Kenya's retail industry with opening of 57th branch. The Star. Accessed on 13 March 2024 at https://www.the-star.co.ke/

Tuzla, H., & Cobb, S. (2012). Evaluation of retail store location alternatives for investment decisions using the Delphi technique and geographic information systems. *International Business: Research, Teaching and Practice*, *6*(2).

Ugoh, C. I., Onyia, C. T., Okoh, J. E., & Omoruyi, P. O. (2022). IS SINGLE FORECAST METHOD BETTER THAN COMBINED FORECAST METHOD?.

Wang, T. K., Gu, X., Li, K., & Chen, J. H. (2021). Competitive location selection of a commercial center based on the vitality of commercial districts and residential emotion. *Journal of Urban Planning and Development*, *147*(1), 04021001.

Wang, Y., Li, S., Zhang, X., Jiang, D., Hao, M., & Zhou, R. (2020). Site selection of digital signage in Beijing: A combination of machine learning and an empirical approach. *ISPRS International Journal of Geo-Information*, *9*(4), 217.

WDG Consulting(2024). Corporate Site Selection Consultants**.** WDG Consulting. Accessed on 26 February 2024 at https://www.wdgconsulting.com/

Widaningrum, D. L., Surjandari, I., & Sudiana, D. (2020). Discovering spatial patterns of fast-food restaurants in Jakarta, Indonesia. *Journal of Industrial and Production Engineering*, *37*(8), 403-421.

Yee, H. J., Ting, C. Y., & Ho, C. C. (2019). Retail site selection using machine learning algorithms. *Int. J. Recent Technol. Eng.(IJRTE)*, *8*, 2422-2431.

Zhao, J., Zong, B., & Wu, L. (2023). Site Selection Prediction for Coffee Shops Based on Multi-Source Space Data Using Machine Learning Techniques. *ISPRS International Journal of Geo-Information*, *12*(8), 329.

8. Appendix

Below are the extra materials that were relevant to my study but could not fit in the main text of my write up

1. Random Forest Algorithm

a) Classification

```
#RF CLASSIFICATION
#read data
data1 <- read.csv("C:\\GIS\\QuickMart Project\\Records\\Classification data.csv")

#explore data
str(data1)
dim(data1)
summary(data1)

#making supermarket as dependent variable
data1$X20 <- as.factor(data1$X20)
table(data1$X20)

#creating training and testing datasets
set.seed(123)
ind <- sample(2, nrow(data1), replace = TRUE, prob = c(0.7, 0.3))
train <- data1[ind==1,]
test <- data1[ind==2,]

#Getting variable of importance
library(randomForest)
set.seed(222)
rf<-randomForest(X20~.,data=train)
print(rf)

#Plot of significance of variables
#imp = importance(rf)
#varImpPlot(rf)

#ACCURACY OF BINARY CLASSIFICATION
#Predict - Test
predictions <- predict(rf, newdata = test)

# Model evaluation metrics
#install.packages("caret")
library(caret)
conf_matrix <- confusionMatrix(predictions, test$X20)
conf_matrix
accuracy <- conf_matrix$overall["Accuracy"]
precision <- conf_matrix$byClass["Precision"]
recall <- conf_matrix$byClass["Recall"]
f1 <- conf_matrix$byClass["F1"]

# Print or use the metrics as needed
cat("Accuracy:", accuracy, "\n")
```

```
cat("Precision:", precision, "\n")
cat("Recall:", recall, "\n")
cat("F1 Score:", f1, "\n")

#PREDICT - Training
predictions <- predict(rf, newdata = train)

# Model evaluation metrics
#install.packages("caret")
library(caret)
conf_matrix1 <- confusionMatrix(predictions, train$X20)
conf_matrix1
accuracy1 <- conf_matrix1$overall["Accuracy"]
precision1 <- conf_matrix1$byClass["Precision"]
recall1 <- conf_matrix1$byClass["Recall"]
f11 <- conf_matrix1$byClass["F1"]

# Print or use the metrics as needed
cat("Accuracy:", accuracy1, "\n")
cat("Precision:", precision1, "\n")
cat("Recall:", recall1, "\n")
cat("F1 Score:", f11, "\n")




# Getting variable importance
importance_values <- importance(rf)
varImpPlot(rf)
# Print the variable importance
#print(importance_values)

# Save the variable importance results as a CSV file
#write.csv(importance_values, file = "variable_importance_results.csv", row.names =
FALSE)
```

b) Regression

```
library(randomForest)
data3 = read.csv("C:\\GIS\\QuickMart Project\\Records\\centres\\road_kde.csv")

str(data3)
set.seed(4543)
rf.fit <- randomForest(Supermarkets ~ ., data=data3, ntree=1000,
                       keep.forest=TRUE, importance=TRUE)
rf.fit

# Make predictions on the same dataset
predictions <- predict(rf.fit, newdata = data3)
```

```
pa <- data.frame(name = data3$Name, actual = data3$Supermarkets, predicted =
predictions)
#difference btw p and a
write.csv(pa, file = "pa_diff2.csv", row.names = TRUE)
# Model evaluation metrics
mse <- mean((data3$X20 - predictions)^2)
rmse <- sqrt(mse)
mae <- mean(abs(data3$X20 - predictions))

# Print or use the metrics as needed
cat("Mean Squared Error (MSE):", mse, "\n")
cat("Root Mean Squared Error (RMSE):", rmse, "\n")
cat("Mean Absolute Error (MAE):", mae, "\n")


#Save importance values
importance_values=importance(rf.fit)
varImpPlot(rf.fit)
#write.csv(importance_values, file = "new regression rf.csv", row.names = TRUE)
```

2. Gradient Descent Algorithm

```
# Load the dataset
data1 <- read.csv("C:\\GIS\\QuickMart Project\\Records\\Before pruning
data\\Regression data.csv")


# Standardize the features
standardize <- function(x) {
  return((x - mean(x)) / sd(x))
}

# Assuming you have features X1 through X19
for (i in 1:19) {
  col_name <- paste0("X", i)
  data1[[col_name]] <- standardize(data1[[col_name]])
}

# dependent and independent variables
X <- data1[, grep("^X", names(data1))][-ncol(data1)]  # Exclude X20
X
Y <- data1$X20
Y
# Standardize the target variable (X20)
#data$X15 <- standardize(data$X15)

# Hypothesis Function for Multiple Linear Regression
hypothesis <- function(theta, X) {
  return(sum(theta * X))
}
```

```r
# Cost Function (Mean Squared Error)
cost_function <- function(theta, X, Y) {
  m <- length(Y)
  h <- hypothesis(theta, X)
  return((1/(2*m)) * sum((h - Y)^2))
}

# Gradient Descent Function for Multiple Linear Regression
gradient_descent_multiple <- function(X, Y, learning_rate, num_iterations) {
  m <- length(Y)
  n <- ncol(X)
  theta <- rep(0, n)  # Initial parameters

  for (iteration in 1:num_iterations) {
    h <- hypothesis(theta, X)

    gradients <- rep(0, n)
    for (j in 1:n) {
      gradients[j] <- (1/m) * sum((h - Y) * X[, j])
    }

    # Update parameters
    theta <- theta - learning_rate * gradients
  }

  return(theta)
}

# Set hyperparameters
learning_rate <- 0.01
num_iterations <- 1000


# Apply gradient descent
theta <- gradient_descent_multiple(X, Y, learning_rate, num_iterations)

# Create a data frame with variable names and their respective theta
theta_df <- data.frame(Variable = names(X), Theta = theta)

# Write to CSV file
write.csv(theta_df, file = "learned_parameters.csv", row.names = FALSE)

# Print the learned parameters
cat("Learned parameters (theta):\n")
print(theta_df)

# Print the learned parameters
cat("Learned parameters (theta):", theta, "\n")
predictions <- hypothesis(theta, X)

# Mean Squared Error (MSE)
mse <- mean((predictions - Y)^2)

# Root Mean Squared Error (RMSE)
rmse <- sqrt(mse)
```

```r
# Mean Absolute Error (MAE)
mae <- mean(abs(predictions - Y))



# Print the metrics
cat("Mean Squared Error (MSE):", mse, "\n")
cat("Root Mean Squared Error (RMSE):", rmse, "\n")
cat("Mean Absolute Error (MAE):", mae, "\n")
```

3. OLS Algorithm

```r
#OLS Regression
data1 <- read.csv("C:\\GIS\\QuickMart Project\\Records\\After pruning data\\RF PRUNED
DATA.csv")

# Fit linear regression model
model <- lm(X20 ~ ., data = data1)
summary(model)
#write csv
coefficients <- coef(summary(model))
significance <- coefficients[, "Pr(>|t|)"]

# Create a dataframe with variable names and significance values
significance_df <- data.frame(Variable = names(significance), Significance =
significance)

# Write to CSV file
#write.csv(significance_df, file = "OLS pruned results.csv", row.names = TRUE)

# Print the dataframe
print(significance_df)

#evaluation metrics
# Making predictions
predictions <- predict(model, data1)

# Calculate Mean Squared Error (MSE)
mse <- mean((data1$X20 - predictions)^2)

# Calculate Root Mean Squared Error (RMSE)
rmse <- sqrt(mse)

# Calculate Mean Absolute Error (MAE)
mae <- mean(abs(data1$X15 - predictions))

# Print the results
cat("Mean Squared Error (MSE):", mse, "\n")
cat("Root Mean Squared Error (RMSE):", rmse, "\n")
cat("Mean Absolute Error (MAE):", mae, "\n")
```

4. Table displaying site rankings

Rank	Name	Difference	X	Y
1		0.965361	254189.4	9856322
2		0.941531	255124.6	9856420
3		0.854284	255946.4	9856793
4		0.849803	257189.8	9858226
5	Muthaiga Square	0.830418	257078.7	9858237
6		0.825109	257154.7	9858125
7		0.808541	257099.4	9858289
8	Harambee Hall	0.807229	257235.2	9858160
9	The Address	0.804179	257371.8	9857868
10		0.765531	257139.1	9858159
11		0.744392	257328.5	9857845
12	Professional Centre	0.740995	257289.5	9857911
13	Muthaiga Business Centre	0.737559	257243	9858005
14		0.73363	257758.3	9857711
15	Milestone Business Centre	0.715501	257724.3	9857807
16	The Stables Karen	0.708294	257701.2	9857793
17		0.691837	257642.7	9857763
18		0.671786	257565.2	9857789
19		0.659533	257544.9	9857852
20	Good Man Tower	0.65607	257583.4	9857855
21		0.649261	257344	9857721
22	The Arches	0.643654	257493.1	9857930
23	WestEnd Towers	0.642018	257420.6	9857894
24	Euitorial Fidelity centre	0.623338	257608.9	9857868
25	Safaricom House 2	0.619602	257689	9857843
26	Hurlingham Plaza	0.612171	257540.4	9857955
27		0.610142	257620.9	9857805
28	Jumuia Place	0.607557	256984.6	9856335
29		0.600851	257142.6	9860716
30		0.596992	256388.4	9860100
31	Nursery garden	0.596706	254199.2	9857222
32	KUCC RESTAURANT	0.587807	258981.6	9857485
33	Lion Place	0.587679	258913.9	9857568
34	Allianz Plaza	0.5809	258959.5	9857518
35		0.574466	258927.1	9857558
36		0.564368	258973.5	9857509
37	Seasons Airport	0.539555	259045.9	9857495
38		0.533579	259020	9857509
39		0.530399	258933.3	9857536
40	Nyaku House	0.519878	258982	9857535

41	Prof. Nelson Awori Centre	0.517943	258987	9857507
42	Brick Court	0.516399	258898.5	9857575
43	JF Centre	0.515541	258945.8	9857526
44	Jamii Towers	0.512456	259012.1	9857495
45	Unipen Flats - Block B	0.512435	258999.4	9857482
46	Dunhill Towers	0.511459	257848.4	9857939
47	Oracle Tower	0.503098	257786.3	9857988
48	Parkfield Place	0.499228	258039.1	9857926
49	Golden Ivy	0.499155	258022.2	9857938
50	Fortis Suites	0.490712	257922	9857995
51	Riara Corporate Suites	0.482019	257997.9	9857981
52	Kenya Association Of Manufacturers	0.477249	258011.4	9858034
53		0.473793	257984.4	9858053
54	UNGA House	0.47316	257676.8	9858011
55	Devan Plaza	0.471425	257709.8	9858164
56	Corporate Business Centre	0.455837	257619.9	9858121
57	Sanlam Tower	0.45557	257623.1	9858051
58	General Accident House	0.449685	257578.8	9858202
59	Riverside Greenpark	0.441042	257640.8	9858154
60	Baraza Cafe	0.435176	257540.4	9858166
61	Grenadier Tower	0.432419	257931.1	9858096
62	Upperhill Gardens apartments	0.430636	257687.2	9857910
63	Real Towers	0.427684	257543.2	9858343
64		0.424259	257569.7	9858288
65		0.421091	257525.4	9858309
66		0.413408	257519.4	9858257
67	Renaissance Corporate Park	0.413002	257602.7	9858742
68	Le'Mac	0.40055	256317.9	9856834
69	Twinster Towers	0.396806	255904.9	9856673
70	Vitendi	0.396647	256025.5	9856643
71	Valley View Plaza	0.39404	256222.1	9857247
72		0.38797	257042.9	9859178
73	The Aerlink	0.382971	257232.6	9859010
74	KWFT Centre	0.377941	258722.9	9859647
75	Corner Plaza	0.377102	255771.4	9861112
76	PWC Tower	0.372963	255163.1	9860560
77	KMA Centre	0.372819	255681.2	9860403
78		0.368111	255755.1	9860797
79	Acorn House	0.366595	255358.4	9860267
80	Tokyo Restaurant	0.36519	255430.5	9860200
81	Pangani Auction Centre	0.362936	255463.4	9860240
82	Fortis Tower	0.358068	255639.3	9860103
83	Geomaps Centre	0.351936	255630	9860457

84		0.350185	255241.6	9860407
85		0.346403	255915.3	9860226
86		0.341349	257848.9	9857770
87	Occidental Plaza	0.330662	257999.7	9857754
88	Sky Park	0.328226	253568.7	9860742
89	Greenspan Shopping Mall	0.328148	255278.8	9860100
90	Simba Auto Works	0.322718	255069.1	9856706
91	Plessey House	0.321587	257970.8	9860539
92	Wanandege Plaza	0.31564	257943.3	9860557
93	Kenrail Towers	0.313194	258055.7	9856010
94		0.310738	257804.3	9857845
95		0.307869	256674.5	9857706
96	Whitefield Palace	0.299792	259177.8	9857476
97	Apollo Center	0.29872	260344.2	9856269
98		0.294126	259989.3	9856071
99	Delta Riverside Office park	0.293061	259978	9856151
100		0.289582	258641.7	9855141
101	Tourism Fund Building	0.286897	258588.8	9855019
102	Tender Care Dental	0.283591	256013.1	9863549
103		0.282829	255494.6	9863646
104		0.280996	250163.5	9856196
105	Riverside Mews Building	0.27718	253086.9	9856287
106	Centenary House	0.275385	260491.4	9854193
107	Landmark Plaza	0.271156	260522.3	9854245
108		0.270592	260481	9854206
109		0.269755	260296.2	9854697
110	Bishan Plaza	0.269505	260379.8	9854662
111	Maksons Plaza	0.267889	260443.2	9854636
112	FCB Mihrab	0.26606	260393.6	9854604
113	Jumuia Place	0.264033	260371.3	9854607
114	Centro House	0.257157	260466.7	9854589
115		0.253866	261172	9854379
116	Old Mutual	0.252013	256725.3	9855753
117		0.248809	262594	9860707
118		0.247821	265887.1	9865024
119		0.247653	265886.2	9865002
120		0.24695	265196.3	9864515
121	Madonna House	0.246913	265008	9864192
122		0.245042	256761	9864848
123	Bemuda Plaza	0.243154	253330.3	9863045
124	Kalamu House	0.239063	264846	9864242
125	CVS Plaza	0.237006	264977	9864214
126	Upper Hill Medical Centre	0.234442	264988.3	9864184

127	Kose Heights, Hurlingham	0.224901	266741.1	9863191
128	Waumini House	0.224544	247947	9852218
129		0.219305	254072.2	9851730
130	Unipen Flats - Block A	0.214977	253960.3	9852287
131	Kaka House	0.206593	253585.7	9852632
132	Sound Plaza	0.204467	253578	9852570
133	Senteu Plaza	0.200444	254382.4	9851666
134		0.199908	254402.4	9851686
135		0.199416	253201.9	9852673
136	Aga Khan Sports Club	0.197716	253202.5	9852686
137		0.193365	262015.8	9861336
138		0.190194	262019.9	9861315
139	Amee Arcade Building	0.188451	262024.5	9861295
140	The Courtyard	0.185968	267685.9	9855129
141		0.182089	265829	9865035
142	Laiboni Centre	0.177271	257265.2	9857891
143	Rainbow Tower	0.176235	257301.3	9857891
144		0.175211	257372.6	9857941
145		0.1715	256691.8	9856339
146	Arnold Plaza	0.170296	256460.9	9856362
147	Tulip Tower	0.165301	257920.5	9857881
148	Mayfair Suites	0.16096	254689.4	9860571
149	MultiChoice Kenya	0.15987	253315.8	9860632
150	Kipro Centre	0.155887	256321.2	9860230
151		0.147997	255693.7	9860702
152		0.147016	255250.1	9860510
153	Westlands Square	0.143199	260320.3	9854687
154	Elysee Plaza	0.140935	260302.1	9854694
155		0.140814	260331.7	9854682
156	Park Suites	0.136518	260343.5	9854678
157	Kenya Institute of Mass communication Ladies Hostels	0.134392	260337.6	9854680
158	KIMC Femaile Hostels	0.128321	260313.9	9854690
159	The Crescent	0.122926	260307.9	9854692
160		0.119343	260290.3	9854699
161	Bandari Plaza	0.117905	260326.1	9854685
162	Westwood Office Building	0.115868	255323.6	9859422
163	ACK Gardens Annex	0.113761	258743.1	9855112
164		0.08351	258560.1	9854910
165	G.V. Plaza	0.080536	258529.5	9854980
166	Aga Khan Sports Club Garden	0.08027	258522.1	9855070
167	Sclaters House	0.071961	258922.6	9855250
168	Amber House	0.07022	259043.4	9855446
169		0.069193	259040.6	9855419

170	Mpaka House	0.068578	266867.7	9854117
171	Valley View Plaza	0.067285	260750.7	9853659
172	KIMC Film Complex	0.063581	260790.4	9853694
173		0.062746	260180.3	9854890
174		0.057581	254177.8	9857755
175	9 West	0.053513	257964.6	9857132
176		0.052066	253775.8	9856781
177	Titan	0.051479	256474.5	9857070
178	KIMC Administration Building	0.047979	266434.7	9857449
179	Krishna Centre	0.047502	256545.2	9859422
180	Victoria Courts	0.035179	255813.1	9860328
181	One Africa Place	0.020696	254423.9	9857389
182	Fehda Plaza	0.015516	254379.2	9857383
183	Adlife Plaza	0.013282	246034.7	9853933
184	Sifa Towers	0.012146	254408.6	9857625
185	Parklands Plaza	0.0106	254483.4	9857393
186	Victoria Plaza	0.007395	257883.5	9857913
187	Aly's Centre	0.002212	254483.4	9857332
188		0.001541	254889.9	9860615
189	Saachi Plaza	-0.01841	254937.5	9860605
190	Cargen Building	-0.01899	258532.1	9860841
191		-0.02102	258564.3	9860806
192		-0.02205	253999.2	9857569
193	Emeli	-0.02493	252727	9857034
194	College House	-0.0253	252772.2	9857046
195	Ojijo Plaza	-0.03576	253307.8	9857154
196		-0.03897	269808.6	9869183
197	Muthithi Place	-0.04234	251323.2	9856459
198	KUSCCO Centre	-0.04619	253352.9	9857402
199	Kam Place	-0.04793	255606.8	9860655
200	53 PARK	-0.04873	259825.3	9863975
201	Forest Road Plaza	-0.05029	255144.8	9859538
202	Vision Towers	-0.05053	255233	9859489
203	Westlands Arcade	-0.05271	255383.2	9859890
204	TRV Office Tower	-0.07196	255675.8	9859908
205	Tender Care Dental	-0.07976	255842.7	9860085
206		-0.09614	256190.5	9859693
207	Centre Point Building	-0.10177	255903.4	9860318
208	Muthurua Police Station	-0.10686	255854.2	9860327
209	Mitsumi Business Park	-0.10951	255685.7	9860273
210	Pelican Signs Building	-0.12889	255638.4	9860317
211	Meky Place	-0.13464	257061.3	9859555
212	53 PARK	-0.13578	255676.4	9860525

213	53 PARK	-0.13578	252096.5	9857025
214	53 PARK	-0.13578	246045.3	9853830
215	Kamirembe Place	-0.14448	253701.1	9857348
216		-0.15886	253788.2	9857334
217	Antarc Office & Home Furniture Solutions	-0.1685	256364.7	9855835
218	Wu Yi Plaza	-0.1695	255824.6	9856776
219	Rosami Court	-0.17122	256153.7	9856382
220	Amani Plaza	-0.17749	253827.6	9856300
221	Bishop Magua Centre	-0.17897	255086.3	9856687
222	Mustek East Africa	-0.17981	255094.2	9856723
223	The Westery	-0.18519	254993.8	9856739
224	Fruity Fruits Plaza	-0.18524	254955.9	9856756
225	53 PARK	-0.19664	254093	9856962
226	Applewood Adams Commercial Center	-0.20005	255032.8	9856819
227		-0.21745	252799	9856355
228	Saachi Plaza	-0.21931	254729.8	9857742
229	The Citadel	-0.22424	254212.5	9860511
230		-0.22878	255002.7	9860408
231		-0.22943	254039.7	9860445
232		-0.24025	253490	9860588
233	Avocado Towers	-0.27573	253651.2	9860583
234		-0.28044	255080.7	9860366
235	Finance House	-0.28363	255525.9	9860086
236		-0.2879	255594.8	9860159
237		-0.29756	255458.4	9860878
238	Rattanasi Trust Building	-0.3296	253823.8	9856472
239	Posta Bank Pension Towers	-0.35192	253810.6	9856589
240	Cresent Business Center	-0.35525	253815.7	9856621
241	Jamahiriya House	-0.35914	253801.1	9856651
242		-0.37198	253779.8	9856601
243	Eco Bank Towers	-0.37341	253800.1	9856568
244	West Park Suites	-0.38731	256701.3	9860092
245	Ellis House	-0.40746	252688.6	9856585
246	Chester House	-0.41459	256174.8	9860209
247	Emperor Plaza	-0.41688	257550.2	9858723
248	Paramount Plaza	-0.41694	255982.1	9859935
249	Shirika Coop House	-0.42155	256246.8	9859653
250		-0.42472	255523.2	9860606
251	Mercantile House	-0.42677	257204.9	9857973
252	Nyingo Towers	-0.43036	257187	9858014
253	Optica Building	-0.4378	257858.5	9857857
254	Library	-0.45044	256488.2	9859976
255	Giwa House	-0.45489	255555.2	9860069

256	Twiga Towers	-0.45534	255675.1	9860079
257	Pension Towers	-0.4581	255629.5	9860141
258	Nanak House	-0.46444	255588.6	9860183
259	Eagle House	-0.46827	255618.9	9860259
260	Marshalls	-0.47084	255809.7	9860137
261	Dominon House	-0.47387	257787.5	9857750
262		-0.47882	257756.4	9857822
263	Uchumi	-0.4832	257781.9	9857835
264	Library	-0.48615	255757.3	9860055
265		-0.48643	257666.8	9858783
266	Kipande Plaza	-0.49419	259466.1	9860157
267	El-Roi Plaza,	-0.50525	257970.8	9856228
268	Norwhich Union House	-0.51539	258034	9856105
269	Uganda House	-0.51592	256267.6	9859633
270	Jubilee Exchange House	-0.52432	255916.8	9859993
271	Teleposta Towers	-0.53784	254186.4	9857467
272	Hughes Building	-0.55802	254723.3	9857378
273	Uniafric House	-0.56198	254722.7	9857354
274	Finance House	-0.56315	254746.5	9857361
275	Old Mutual Building	-0.56662	254494.7	9857292
276	National Bank	-0.56964	253031.7	9860680
277		-0.57841	256157.9	9853671
278	Odeon	-0.57922	255864.3	9860175
279	Post Bank	-0.58772	255424.6	9860009
280	Library	-0.60004	255421.9	9859963
281	CFC Stanbic Bank	-0.60309	249237.2	9855764
282	Lyric Building	-0.60434	251421.9	9858008
283	Stannbank House	-0.61424	256518.8	9859478
284	Hamilton House	-0.6202	256477.8	9859512
285	Phoenix House	-0.62067	256464.6	9859529
286	Kimathi House	-0.63629	256319.1	9859571
287	International House Ltd	-0.63733	255946.4	9860117
288	Stanbank Building	-0.66531	257381.1	9855713
289		-0.67694	256471.1	9859643
290	Jubilee Insurance Place	-0.68749	257082.9	9859836
291		-0.6992	259920.8	9860476
292	The Bazaar	-0.71871	257976.7	9856141
293	Vedic Building	-0.72212	255901.6	9856713
294	Town House	-0.72271	256892.7	9855898
295	Rehani House	-0.73025	257153.6	9855982
296	Standard Building	-0.73881	256419.1	9860036
297	Revlon Plaza	-0.73908	255733.8	9860132
298		-0.74261	255873.5	9860180

299		-0.74387	255814.9	9860472
300	KenCom Building	-0.75291	254262.4	9859634
301		-0.75616	255390.7	9860036
302		-0.75863	255368	9859924
303	Bihi Tower	-0.76415	257824.1	9857854
304		-0.76579	255532.3	9861340
305	Ufundi Coop House	-0.7708	255321.2	9865475
306		-0.77694	256424.7	9855857
307	JKUAT Towers	-0.77904	267456.7	9855289
308	20th Century Plaza	-0.78096	267419.6	9855269
309	Corner House	-0.7875	267485.3	9855271
310	Sanlam House	-0.78751	267455.7	9855239
311		-0.79122	267536.7	9855219
312		-0.7949	267614.5	9855272
313	Prudential Assurance	-0.79523	267605	9855264
314		-0.79906	267811	9855147
315	Queensway House	-0.81457	270059.3	9852576
316	Gilifan House	-0.82855	270046.9	9852556
317	Kenya Charity Sweeptake House	-0.83876	270054.7	9852613
318		-0.83959	264912.2	9859046
319		-0.87764	254023.2	9857707
320		-0.88421	251744.9	9858228
321		-0.89718	253408.3	9857355
322		-0.93151	253342.5	9857461
323	Transnational Plaza	-0.93722	253446.2	9857336
324	Hamburg House	-1.02032	253375.2	9857342
325		-1.02062	253665.7	9857335
326		-1.13081	253669.7	9857312
327		-1.16364	253660.6	9857357
328		-1.44054	252094.8	9857038
329		-1.57971	269159.5	9869079