Application of Agglomerative Hierarchical Clustering to Classification of Housing Units in Nigeria by Types of Construction Materials

H.B Joseph_Ikughur, S.C Nwaosu and F.E Udoumoh.

Department of Mathematics/ Statistics/ Computer Science, Joseph Sarwuan Tarka University,Makurdi,Nigeria

OPEN ACCESS

****Correspondence:***
H.B Joseph_Ikughur. Department of Mathematics/ Statistics/Computer Science, Joseph Sarwuan Tarka University,Makurdi,Nigeria.

Specialty Section:
This article was submitted to Maths/Computer, a section of NAPAS.

Submitted: *1 May2022*
Accepted: *13 October 2022*
Published: *15 November2022*

Citation:
H.B Joseph_Ikughur, S.C Nwaosu and F.E Udoumoh. (2022). Application of Agglomerative Hierarchical Clustering to Classification of Housing Units in Nigeria by Types of Construction Materials. Nig Annals of Pure & Appl Sci. 5(2):335-352. DOI:10.5281/zenodo.7338445

Publisher:
cPrint, Nig. Ltd
E-mail: *cprintpublisher@gmail.com*

Access Code

http://napas.org.ng

ABSTRACT

Housing and quality of houses can be categorized based on several criteria including cost of constructions, quality of materials, construction quality, location of building among others. In this paper, Agglomerative Hierarchical Clustering (AHC) technique is utilized to classify housing by types of material used in their construction as well as their locations. Single Linkage and the Ward's methods along with Analysis of Variance (ANOVA) were used for analysis. The result showed that three (3), four (4) and four (4) clusters were respectively formed for roof, floor and wall types of materials. Cluster membership was used to identify the locations of this clusters that was formed and materials used at these locations with the help of final cluster center. The Analysis of variance (ANOVA) which validated the results in this paper showed that cluster classification of the roof, floor and wall types in Nigeria by Hierarchical Agglomerative Clustering method was good.

Keywords: Agglomerative Hierarchical Clustering, Classification, Housing, ANOVA

1.0 INTRODUCTION

A cluster population is one with heterogeneous elements or one whose units are heterogeneous in nature. In modeling, Heterogeneity tends to negatively affects the assumption of linear models that suppose that units are homogeneous. Apart from stratification and discriminant analysis, another way out is to group the cluster unit into within groups or between groups using the method of cluster analysis which is suggested as a practical method for identifying meaningful clusters within samples that may appear homogeneous (Yim and Ramdeen, 2015).Clustering is the task of grouping a set of objects in such a way that objects in the same group are more similar to each other and dissimilar to those in the other groups. According to Allen (2017), a hierarchical model refers to a type of data analysis structure whereby the data are organized in a tree like structure or that employs multilevel modeling. Renchers (2002), stated that Hierarchical methods and other approaches permit us to search for a reasonable solution without having to look at all possible arrangements. Housing on the other hand is an investment having a significant role in the life of individual, local and national economy. According to Bello (2019), housing is a physical structure meant for the provision of shelter or accommodation to its occupants which can be used to protect the occupants from hash effects of weather and other dangerous things that may cause harm to individual lives.

Henilane (2016) describe housing as one of the most important life components that gives shelter, safety and warmth, as well as finding a place of rest. In the opinion of the author, the need for housing is not only one of the basic human need, but also indicator of living standard of the population. It is worthy of note that the important criteria for housing includes its comfortability, economical, reasonable maintainability as well as architecturally expensive etc. (Henilane,2015a).

It is on this note that Grimes and Orville (1976) explained that housing has been accorded with a physical phenomenon mostly related with cost of construction that largely depend on type of materials used, construction quality and housing standards. Thus, variables such as cost of constructions, quality of materials, construction quality, location of building among other specifications are worthy of note. This is because, in planning housing policies and development, information realized from these indicators will help in reshaping the policies that will bring about development of change in housing sector. Therefore, this study performs classification of housing in Nigeria by means of hierarchical agglomerative clustering technique considering materials used in building construction as well as location of the buildings.

2.0 LITERATURE REVIEW

According to Stafford (1978), the concept of housing is generally defined for statistical purposes as dwelling units (Housing unit occupied separately by households) comprising a great variety of quantities and qualities. Aroni (1982) and Achuenu (2002), pointed out that housing should be a home, a resting place with fundamental purpose of a secured, rewarding, happy or at least a livable space. In the context of socio-cultural functionality, housing is viewed as an area for recreation and identification (Gallent *et al*; 2004) and can be regarded as psychological identity, a foundation for security and self-respect (Aroni, 1982) societal support (Johnson, 2006) and the setting for the formation of social relationships (Amole, 1997).

A house can be seen as a dwelling place for human habitation whether a crude hut or an elaborate mansion, and whatsoever its degree of intrinsic architectural interest, a house as a matter of fact, provides shelter and acts as a focal point for day-to-day living (Ahsen and Gulcin, 2005). Okupe (2002) postulated housing as a strategic asset to man, irrespective of his/her socio-economic status, colour or creed and as such there lies a passion and emotional attachment to housing in African traditional setting. Housing is a fundamental product of every human effort irrespective of his or her financial standing. However, the level of housing production in Nigeria despite the number of housing programmes and policies is still at its lowest ebb. Provision of shelter therefore is that passive and primary function of housing while its secondary function is the creation of an environment that is best suited to the way of life of a people. Furthermore, housing symbolizes the social status of the family to both the wider community and in the nuclear family setting itself. The quality and quantity of housing stock is a reliable barometer of measuring the technology, culture and above all, civilization level of any nation. A house with all its necessary physical attributes must have a rich set of evolving cultural, demographic and psychological meanings attached to it to be called a home.

Basically, Housing is classified by housing types size, amenities, location, group of population living in the housing, type of ownership rights, construction period of the housing energy efficiency, includes, construction materials used in the interior and exterior with and by other features. Central Statistical Bureau of Latvia (2005) cited in Henilane (2016) present a tabulation of criteria for housing classification as shown in Table1

Yim and Ramdeem (2015) looked at cluster analysis as a class of data reduction methods used for sorting cases, observations, or variables of a given data set into homogeneous groups that differ from each other. They also posited that the purpose of cluster analysis is to discover a system of organizing observations where members of the groups share common characteristics. Furthermore, cluster analysis is a class of techniques that classifies cases into groups that are relatively homogeneous *within* themselves and relatively heterogeneous between each other (Landau and Ster, 2010), Clustering technique includes the following: Hierarchical algorithms, Partitioning algorithms and Bayesian algorithms.

Table1: Classification of Housing

Types of housing classification	Characteristics
By housing type	Room in the apartment, Apartment in Multi-apartment residual building or non-residential building Multi- apartment residential building, Family house Other
By housing size	One room, One-Room apartment, Two-Room apartment, Three-Room apartment, Family house Other
By housing amenities	Housing with all amenities. Housing with part of amenities. Housing without amenities.
By housing location	Housing in a city, Housing in rural territory
By group of population living in the housing	Any resident, Persons with low –income or other social group at risk
By type of housing ownership rights	State-owned housing, Municipality-owned housing Natural persons owned housing; Legal persons owned housing other
By construction period of the housing	Housing build before World War11, housing Built from 1945 to 1990, housing built from 1990 until now
By energy efficiency indicators of housing	Minimum regulatory Energy performance level allowed for new buildings Minimum regulatory energy performance level allowed for reconstructed or renovated buildings almost zero energy consumption housing, other
By construction materials used in the exterior wall of the housing	Brick wall, Wood, Brick/Panel, Reinforced concrete/Concrete, Lightweight Concrete, Wood/masonry, other.

Source: *Central Statistical Bureau of Latvia (2005) cited in the work of Henilane(2016)*

Clustering on the other hand, is a technique for grouping items or objects with similar **or** dissimilar group of clusters. Renchers (2002) considered clustering as a classification, pattern recognition (specifically unsupervised learning) and numerical taxonomy. Cluster analysis technique have been extensively applied to data in many fields of studies which includes Medicine, psychology, geology, criminology, anthropology, geology and etc. whenever researchers are interested in clustering analysis, the target is to find an optimal grouping for which the observations or objects within each cluster are similar but are dissimilar to each other.

Hierarchical algorithm can either be agglomerative algorithm and divisive algorithm. Agglomerative algorithm begins with each element as a separate cluster and merge them into successively larger clusters while divisive algorithm begins with the whole set and proceed to divide it into successively smaller clusters.

In hierarchical clustering, we typically start with n clusters, one for each observation, and end with a single cluster containing all n observations. At each step, an observation or a cluster of observation is absorbed into another cluster, this process can be revised by starting with the single cluster containing *n* observations and end with n clusters of a single item each. However, in partitioning clustering, the observations are divided into *g* cluster which is done by starting with an initial partitioning or with cluster centers and then reallocating the observations according to some optimality criterion.

Application of Cluster Analysis Techniques.

The work of Stigler and Becker (1977); showed that neither housing preferences nor housing demand are homogenous and suggested that heterogeneous customers can be grouped based on similarities they share: - basic characteristics needs, preference, altitude

Yang and Miao (2014) in their study applied cluster analysis to classify modern service industries in China. Based on statistical standards, they classified the modern service industries into nine (9) categories.

Hepsen and Vatansever (2012) in the study of Turkish Residential Market identified clustering as potentially important contribution to real estate portfolio analysis. They utilized hierarchical clustering algorithm to study rental returns for seventy-one metropolitan residential markets by developing homogenous grouping for real estate portfolios. Their findings showed 3-cluster partition of the districts namely districts with lowest rental returns, relatively higher rental returns and highest returns.

Shah *et al.* (2013) utilized the K-means and Hierarchical-clustering techniques to assign sense to movies which gives the observer an idea about the generic theme of the movies. This is because, of the continuously increasing number of movies produced over the years and the similarities observed among them. In order to reduce the number of clusters, they classified the 27 identified genres by IMDB (International Movie Data Base) into five groups using hierarchical clustering method. The genres are comedies, action, documentary, drama or horror based on the measurable qualities of the musical score of the movie.

Soltani and Modarres (2006) in their study of annual rainfall in Iran used hierarchical clustering analysis to classify annual rainfall over Iran into spatial groups and it was found that hierarchical clustering analysis are more suitable for classification over spatial groups in Iran. They also observed that the derived clusters and geographical conditions are very well matched with each other and that the ward's methods matched very well with result of the factors influencing rainfall in Iran.

Liao *et al.* (2016) in their study of end-stage renal disease patients who initiated hemodialysis used hierarchical clustering analysis and K-means cluster analysis with either flexible Beta or Ward's methods and found that the K-means cluster analysis method was more suitable in healthcare claims data with highly skewed cost information when taking into account both change of cost patterns and sample size in the smallest cluster.

Pan *et al.* (2019) applied cluster analysis and discriminant analysis in quality grinding of jadeite Jude. Important results recommended by K-means Clustering and the ANOVA table indicated the validity of the five classifications, hence, ANOVA table is a useful measure of validity.

Abraham *et al.* (1994) utilized clustering techniques to identify structural relationships among United States (U.S.) housing markets and develop a bootstrapping procedure to test whether associations between cities are significant. The method was used to create meaningful "groups" of cities that are useful for purposes of diversification, and for identifying appropriate hedging proxies for city-specific futures instruments. Here the *K*-means algorithm was applied to the 1977-1992 returns to housing price indices in 30 metropolitan U.S. housing markets. It demonstrates strong regional differences in housing price fluctuations as they found out that when three groups are specified, a West Coast group, East Coast group, and a Central U.S. group were formed and when more groups are specified, the West Coast divides into two clusters that are North and South, and Texas cities separate from the Central U.S. group.

In Nigeria, Morekenji *et al.* (2012) grouped housing by physical qualities across Nigeria grouped into three regions namely; East, North and South. The paper utilized principal component Analysis and it showed that the classification was 93% correct.

There exist generally few works that studied housing classification in Nigerian from statistical viewpoint. Works utilizing Agglomerative Hierarchical Clustering and their associated K-means techniques are not clearly in view of this researcher. This study therefore utilizes Agglomerative Hierarchical Clustering techniques in order to classify housing by types of material use in their construction and by the locations of these structures.

3.0 METHODS

In this research paper the methods adopted includes the single linkage method known as the Nearest neighborhood , ward method which uses the sum of squares and ANOVA which is used to validate the results obtained from cluster analysis, the data used in this study is of two dimension consisting of both the materials used and the locations of the clusters and was extracted from the National Bureau of Statistics, Annual Abstract of Statistics 2010, and the Software used for the Analysis is SPSS and Minitab.

Consider the data matrix given as

$$Y = \left(y_{(1)}\ y_{(2)}, \ldots, y_{(p)}\right) \quad (1)$$

where y_i' is the n row observation vector and $\mathbf{y}_{(j)}$ is the p column corresponding to a variable, the task is to group the y_i''s (rows) into g clusters and it may also be of interest to cluster the columns , j=1,2, …, p. The data can be represented in a two-dimensional $n \times n$ table.

3.1 Measure of Similarity/dissimilarities

In order to identify the observation vector that are similar and group them into cluster we utilize index of similarity or index of proximity between each pair of observations of which the distance between each pair of observations is of interest. Consider the Euclidean distance between two vectors

$$X = (x_1, x_2, x_3, \ldots, x_p)' \quad \text{and} \quad Y = (y_1, y_2, y_3, \ldots, y_p)'$$

given as

$$d(\mathbf{X}, \mathbf{Y}) = \sqrt{(X-Y)'(X-Y)} = \sqrt{\sum_{j=1}^{p}(x_j - y_j)_2} \quad (2)$$

Adjusting the variance and covariance among the p-variables, Equation (2) is now modified to be

$$d(\mathbf{X}, \mathbf{Y}) = \sqrt{(X-Y)'S^{-1}(X-Y)} \quad (3)$$

where S is the sample variance covariance matrix and is given as

$$S = \begin{pmatrix} S_{11} & S_{12} & \cdots & S_{1p} \\ S_{21} & S_{22} & \cdots & S_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ S_{p1} & S_{p2} & \cdots & S_{pp} \end{pmatrix}$$

which can be expressed as

$$S = \frac{1}{n-1}[Y'Y - Y'(\tfrac{1}{n}J)Y] = \frac{1}{n-1}Y'[I - \tfrac{1}{n}J]Y$$

$$\text{Such that } I = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}_{p\times p} \quad \text{and} \quad J = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}_{p\times p} \quad (4)$$

In some cases, the Minkowski distance measure is used and this is given as

$$D(\mathbf{X}, \mathbf{Y}) = \left[\sum_{j=1}^{p}|x_j - y_j|^r\right]^{\frac{1}{r}} \quad (5)$$

If $r = 2$, then the Minkowski metric is equivalent to Euclidean metric.

Again, Consider the n- observation vector $y_1, y_2, \cdots y_n$, the n x n matrix denoted as

$$\mathbf{D} = (d_{ij}) \quad (6)$$

where n $d_{ij} = d(y_i, y_j)$ and is given as

$$d(y_i, y_j) = \sqrt{(y_i - y_j)'(y_i - y_j)} \quad (7)$$

taking square of both side we get

$$d^2(Sy_i \cdot y_j) = (y_i - y_j)'(y_i - y_j) \quad (8)$$

which is a measure of dissimilarity and its usually symmetric with diagonal element =0

Renchers (2002) showed that the squared Euclidean distance between

$X = (x_1, x_2, x_3, \dots, x_p)'$ and $Y = (y_1, y_2, y_3, \dots, y_p)'$ is

$$d^2(x_i, y_j = \sum_{j=1}^{p}(x_j - y_j)^2 \quad (9)$$

this measure can be express as

$$d'(x, y) = (v_x - v_y)^2 + p(\bar{x} - \bar{y})^2 + 2v_x v_y(1 - r_{xy}) \quad (10)$$

such that

$$v_x{}^2 = \sum_{j=1}^{p}(x_j - \bar{x})^2 \quad \text{and} \quad \bar{x} = \sum_{j=1}^{p}\frac{x_j}{p}, \quad (11)$$

$$\bar{y} = \sum_{j=1}^{p}\frac{y_j}{p}, \quad v_y{}^2 = \sum_{j=1}^{p}(y_j - \bar{y})^2 \quad (12)$$

The correlation between x and y, denoted as r_{xy} is given as

$$r_{xy} = \frac{\Sigma_{j=1}^{p}(x_j - \bar{x})(y_j - \bar{y})}{\sqrt{\Sigma_{j=1}^{p}(x_j - \bar{x})^2\,\Sigma_{j=1}^{p}(y_j - \bar{y})^2}} \quad (13)$$

3.2 Hierarchical clustering:

Hierarchical clustering approach is the clustering technique that begins with n clusters, one for each observation and end with a single cluster containing all n-observation. It is worthy to note that at each step an observation or cluster of observation is absorbed into another cluster and this process can also be revised by starting with a single cluster containing all n observations and end with n clusters of a single item each. In an attempt to find good clusters, the number of ways of partitioning a set of n items into g clusters is given by

$$N(n, g) = \frac{1}{g!}\sum_{k=1}^{g}\binom{g}{k}(-1)^{g-k}k^n \quad (14)$$

This approach makes it easy to search for a reasonable solution without looking at all possible arrangements. Basically, there are several approaches to measuring distance between clusters that brings about different hierarchical agglomerative methods. In this study, our interest will be on single linkage and wards methods.

3.3 Single Linkage

This is also known as nearest neighbor method, which is the distance between two clusters A and B defined as the minimum distance between a point in A and a point in B and is given as

$$,D(A,B) = min\{d(y_i, y_j), \quad \text{for } y_i \text{ in A and } y_j \text{ in B}\} \tag{15}$$

Where $d(y_i, y_j)$ is the Euclidean distance as in Equation 2or some other distance between the vectors y_i and y_j. At each step in the single linkage method the distance in Equation 15 is obtained for every pair of clusters and the two clusters with smallest distance are merged thereby reducing the number of clusters by 1, This procedure is repeated and the resulting solutions of the hierarchical clustering is displayed graphically with the aid of a tree diagram known as a dendrogram which shows all the steps in the procedure including the distance at which clusters are merged.

3.4 Ward's Method

Wards method is also known as incremental sums of squares methods which uses the within cluster squared distances (Ward (1963) and Wishart (1969) as cited in Renchers (2002)). If AB is the cluster obtained by combining clusters A and B, then the sum of within cluster distance (of the item from the cluster mean vectors) are

obtained by combining clusters A and B, then the sum of within cluster distance (of the item from the cluster mean vectors) are

$$SSE_A = \sum_{i=1}^{n_A} (y_i - \bar{y}_A)'(y_i - \bar{y}_A) \tag{16}$$

$$SSE_B = \sum_{i=1}^{n_B} (y_i - \bar{y}_B)'(y_i - \bar{y}_B) \tag{17}$$

$$SSE_{AB} = \sum_{i=1}^{n_{AB}} (y_i - \bar{y}_{AB})'(y_i - \bar{y}_{AB}). \tag{18}$$

Where

$$\bar{y}_{AB} = \frac{n_A\bar{y}_A + n_B\bar{y}_B}{n_A + n_B}$$

n_A, n_B and $n_{AB} = n_A + n_B$ are the numbers of points in A and B , and AB, respectively. This sum of distance is equivalent to within cluster sums of squares and are denoted by SSE_A SSE_B and SSE_{AB}

. Ward's method joins the two clusters A and B that maximizes the increase in SSE,defined as

$$I_{AB} = SSE_{AB} - (SSE_A + SSE_B) \tag{19}$$

It can be shown that the increase I_{AB} in 19 has the following two equivalent forms

$$I_{AB} = n_A(\bar{y}_A - \bar{y}_{AB})'(\bar{y}_A - \bar{y}_{AB}) = n_B(\bar{y}_B - \bar{y}_{AB})'(\bar{y}_B - \bar{y}_{AB})$$
$$= \frac{n_A n_B}{n_A n_B}(\bar{y}_A - \bar{y}_B)'(\bar{y}_A - \bar{y}_B) \qquad (20)$$

By Equation 20, minimizing the increase in SSE is equivalent to minimizing the between cluster distance if A consists only y_i and B consists only y_j, then SSE_A and SSE_B are zero as in Equation 19 and Equation 20 and this reduces to

$$I_{ij} = SSE_{AB} = \frac{1}{2}(y_i - y_j)'(y_i - y_j) = \frac{1}{2}d^2(y_i, y_j) \qquad 21$$

Writing the

$\frac{n_A n_B}{n_A + n_B}$ in 20 as $\frac{n_A n_B}{n_A + n_B} = \left(\frac{1}{1/n_A + 1/n_B},\right)$ as n_A and n_B increase

increase, $\frac{n_A n_B}{n_A + n_B}$ increases. Ward's methods have higher likelihood of joining smaller cluster or clusters of equal size.

3.5 Agglomeration Schedule

This is a table that shows which variables or clusters of variables are paired together at different stages of a cluster analysis (Cramer, 2003). In agglomeration schedule, cases with the smallest distance are combined to form a new cluster by joining them at each stage.

3.6 Cluster Membership

A cluster is a network collection of nodes. The final cluster centers are computed as the mean for each variable within each final cluster. The final cluster centers reflect the characteristics of the typical case for each cluster.

3.7 Cluster Validity

Validating the outcomes of hierarchical agglomeration clusters firstly requires the correct determination of the number of clusters. The number of clusters is determined at the point in the scree plot where the elbow begin to rise using information on the agglomeration schedule. Secondly, the ANOVA Table is used to determine if the clusters are significantly different with large F-Statistics or P-value <0.05. Thus, ANOVA helps us to known which of the variables that contribute the most to the cluster solution. Variables with large F values provides the greatest separation between clusters.

4.0 RESULTS

In this section, the results obtained in this study were presented on Table 2 to Table 14 for the materials used. Tables 2, 3 and 4 shows results on proximity matrix for floor types, roof types and wall types while Tables 5,6,7 shows results on agglomeration schedule respectively for roof types, floor types, and wall types Table 8 shows result on cluster membership for roof types, floor types and wall types respectively, in the same way Tables 9 -11 shows the final cluster centers for the materials used and Tables 12- 14 shows results for ANOVA respectively for roof types, floor types and wall types

5.0 DISCUSSION OF RESULTS

4.1 Results on Proximity Matrix

Table2: Proximity Matrix for Roof Types

Case	TPR	WB	EMM	CMZS	SA	CC	RT	OTHERS
TPR	0	2.85E+11	3.74E+11	1.81E+14	1.69E+12	6.70E+11	6.99E+11	7.49E+11
WB	2.85E+11	0	8.73E+10	1.82E+14	1.17E+12	2.40E+11	1.97E+11	2.29E+11
EMMB	3.74E+11	8.73E+10	0	1.82E+14	1.52E+12	5.02E+11	4.36E+11	4.62E+11
CMZS	1.81E+14	1.82E+14	1.82E+14	0	1.82E+14	1.82E+14	1.83E+14	1.83E+14
SA	1.69E+12	1.17E+12	1.52E+12	1.82E+14	0	5.36E+11	1.11E+12	1.24E+12
CC	6.70E+11	2.40E+11	5.02E+11	1.82E+14	5.36E+11	0	1.37E+11	1.96E+11
RT	6.99E+11	1.97E+11	4.36E+11	1.83E+14	1.11E+12	1.37E+11	0	6.05E+09
Others	7.49E+11	2.29E+11	4.62E+11	1.83E+14	1.24E+12	1.96E+11	6.05E+09	0

Table3 : Proximity Matrix for Floor Types

Case	EMB	WB	CC	STONE	BB	VT	CMT	TAR	OTHERS
EMB	0	2.73E+12	6.35E+12	3.46E+12	3.57E+12	3.59E+12	3.55E+12	3.62E+12	3.64E+12
WB	2.73E+12	0	8.74E+12	5.81E+10	7.21E+10	7.54E+10	7.14E+10	7.90E+10	8.05E+10
CC	6.35E+12	8.74E+12	0	9.43E+12	9.55E+12	9.34E+12	9.24E+12	9.35E+12	9.65E+12
STONE	3.46E+12	5.81E+10	9.43E+12	0	3.01E+09	5.25E+09	5.71E+09	6.41E+09	4.17E+09
BB	3.57E+12	7.21E+10	9.55E+12	3.01E+09	0	3.04E+09	4.39E+09	3.93E+09	5.31E+08
VT	3.59E+12	7.54E+10	9.34E+12	5.25E+09	3.04E+09	0	4.56E+08	7.50E+08	3.60E+09
CMT	3.55E+12	7.14E+10	9.24E+12	5.71E+09	4.39E+09	4.56E+08	0	1.43E+09	5.55E+09
TAR	3.62E+12	7.90E+10	9.35E+12	6.41E+09	3.93E+09	7.50E+08	1.43E+09	0	4.15E+09
OTHERS	3.64E+12	8.05E+10	9.65E+12	4.17E+09	5.31E+08	3.60E+09	5.55E+09	4.15E+09	0

Table4 : Proximity Matrix for Wall Types

Case	RM	WB	STONE	CBB	MZ	OTHERS
RM	0	9.73E+12	1.09E+13	1.19E+13	1.05E+13	1.11E+13
WB	9.73E+12	0	9.36E+10	8.32E+12	7.09E+10	1.18E+11
STONE	1.09E+13	9.36E+10	0	9.13E+12	1.89E+10	5.51E+09
CBB	1.19E+13	8.32E+12	9.13E+12	0	8.69E+12	9.22E+12
MZ	1.05E+13	7.09E+10	1.89E+10	8.69E+12	0	2.18E+10
OTHERS	1.11E+13	1.18E+11	5.51E+09	9.22E+12	2.18E+10	0

4.2 Results on Agglomeration Schedules

Table5: Agglomeration Schedule for Roof Types

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	7	8	3.026E9	0	0	3
2	2	3	4.667E10	0	0	4
3	6	7	1.568E11	0	1	5
4	1	2	3.620E11	0	2	5
5	1	6	8.785E11	4	3	6
6	1	5	1.791E12	5	0	7
7	1	4	1.611E14	6	0	0

Table6: Agglomeration Schedule for Floor Types

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	6	7	2.28E+08	0	0	3
2	5	9	4.94E+08	0	0	4
3	6	8	1.14E+09	1	0	5
4	4	5	3.45E+09	0	2	5
5	4	6	8.73E+09	4	3	6
6	2	4	6.98E+10	0	5	7
7	1	2	3.08E+12	0	6	8
8	1	3	1.07E+13	7	0	0

Table7: Agglomeration Schedule for Wall Types

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	3	6	2.75E+09	0	0	2
2	3	5	1.54E+10	1	0	3
3	2	3	8.21E+10	0	2	5
4	1	4	6.04E+12	0	0	5
5	1	2	1.50E+13	4	3	0

4.3 Results on Cluster Membership

Table 8: Cluster Membership for (a) Roof Type (b) Floor Type and (c) Wall Type

		(a) Roof Type		(b) Floor Type		(c) Wall Type	
Case No	State	Cluster	Distance	Cluster	Distance	Cluster	Distance
1	Abia	3	162512.06	4	157751.222	2	123197.09
2	Adamawa	3	190757.7	1	72991.602	3	57249.072
3	Akibom	3	160837.26	4	55549.595	2	65328.576
4	Anam	3	222587.96	4	46114.447	2	62446.5
5	Bauchi	3	260950.22	1	240405.621	3	149750.26
6	Bayelsa	3	359907.52	1	156942.315	3	277230.6
7	Benue	3	199149.57	1	216114.024	3	121650.75
8	Bornu	3	143816.08	1	190040.862	3	58002.232
9	Crrivers	3	111693.14	1	152406.203	3	204077.85
10	Delta	3	253773.82	4	40283.074	1	0
11	Ebonyi	3	150767.86	1	63032.421	3	131897.64
12	Edo	3	227508.31	4	99113.317	2	52219.681
13	Ekiti	3	176484.98	1	219360.35	2	186075.01
14	Enugu	3	169841.09	4	100519.77	2	102160.93
15	Gombe	3	208643.58	1	77770.851	3	157599.12
16	Imo	3	251586.43	4	111299.97	2	75414.297
17	Jigawa	3	333229.5	1	234359.008	3	203047.45
18	Kaduna	3	180365.74	4	215627.605	3	291173.84
19	Kano	3	389068.15	2	163372.463	3	594905.82
20	Katsina	3	373238.68	2	163372.463	3	311874.91
21	Kebbi	3	104956.57	1	130241.896	3	99820.469
22	Kogi	3	192633.92	4	161824.44	2	144627.02
23	Kwara	3	349015.4	1	179808.233	2	236240.15
24	Lagos	2	0	3	0	4	0
25	Nasarawa	3	138992.6	1	143954.158	3	278802.28
26	Niger	3	58032.462	1	173015.201	3	125737.87
27	Ogun	3	313515.99	4	111900.902	2	169121.75
28	Ondo	3	125580.14	4	108529.343	2	81431.287
29	Osun	3	248587.74	4	88966.888	2	26468.843
30	Oyo	3	275247.2	4	271659.554	2	345137.14
31	Plateau	3	393969.52	1	90898.437	3	75125.094
32	River	3	199890.98	4	228813.235	2	248601.5
33	Sokoto	3	269118.07	1	171113.837	3	114422.55
34	Taraba	3	282159.81	1	77840.141	3	144715.58
35	Yobe	3	263885.03	1	153763.096	3	176322.15
36	Zamfara	3	201591.65	1	135640.598	3	120817.9
37	Abuja	1	0	1	223054.993	3	319958.53

4.4 Results on Final Cluster Membership

Table 9: Final cluster Centers for Roof Types

	Cluster		
	1	2	3
TPLR	9179	29955	126209
WB	4249	54876	64525
EMMB	7469	14578	76220
CMZS	**13328987**	218282	**364729**
SA	21616	**990871**	53590
CC	16738	347111	42852
RT	3433	55055	12603
Others	1656.00	11193.00	4097.80

KEY: ***TPLR*** *= Thatch/Palm leaves/Rafia,* ***WB*** *= Wood baboo,* ***EMMB*** *= Earth mud/ Mudbricks,* ***CMZS*** *= Corrugated Metal zinc Sheet,* ***SA*** *= State Asbestos,* ***CC*** *= Cement Concrete* ***RT*** *= Roofing Tiles*

Table 10: Final Cluster Centers for Floor Types

	Cluster			
	1	2	3	4
EMB	265074	777914	76030	204399
WB	39095	118563	43278	27052
CC	209572	380088	1863080	580852
STONE	9699	17213	17735	8751
BB	4174	11874	7943	6084
VT	3152	7843	54163	8985
CMT	4437	10677	59393	12927
TAR	2200	5760	66972	6645
OTHERS	2476	4897	7248	1918

KEY: ***EMB*** *= Earth mud bricks,* ***WB*** *= Wood Baboo,* ***CC*** *= Cement Concrete,* ***BB*** *= Burnt Bricks,* ***VT*** *= Vinyl Tiles,* ***CMT*** *= Ceramic Mable Tiles,* ***TAR*** *=* ***Terrazzo***

Table11: Final Cluster Centers for Wall Types

	Cluster			
	1	2	3	4
RM	2644266	180401	381434	51520
WB	50672	29034	65376	59485
STONE	8030	7610	16363	15963
CBB	526871	530711	168236	2015697
MZ	32676	24281	28935	31670
OTHERS	7597	5608	8292	11507

KEY: ***RM*** *= Reedmud,* ***WB*** *= Wood Baboo,* ***CBB*** *= Cement Blocks/Bricks,* ***MZ****= Metal Zinc*

4.5 Results of ANOVA

Table13: ANOVA for Clusters on Roof Types

	Cluster		Error		F	Sig.
	Mean Square	Df	Mean Square	Df		
TPR	1.09E+10	2	6.90E+09	34	1.574	0.222
WB	1.80E+09	2	3.12E+09	34	0.577	0.567
EMM	4.03E+09	2	8.54E+09	34	0.473	0.627
CMZS	8.18E+13	2	3.49E+10	34	2344.1	0
SA	4.29E+11	2	5.88E+09	34	72.94	0
CC	4.56E+10	2	9.06E+08	34	50.287	0
RT	9.28E+08	2	5.19E+07	34	17.887	0
OTHERS	2.79E+07	2	5674065.87	34	4.91	0.013

KEY: ***TPLR*** *= Thatch/Palm leaves/Rafia,* ***WB*** *= Wood baboo,* ***EMMB*** *= Earth mud/ Mudbricks,* ***CMZS*** *= Corrugated Metal zinc Sheet,* ***SA*** *= State Asbestos,* ***CC*** *= Cement Concrete* ***RT*** *= Roofing Tile*

Table14: ANOVA for Clusters on Floor Types

	Cluster		Error		F	Sig.
	Mean Square	Df	Mean Square	Df		
EMB	2.04E+11	3	1.49E+10	33	13.685	0
WB	4.89E+09	3	5.95E+08	33	8.224	0
CC	1.11E+12	3	1.13E+10	33	97.893	0
STONES	6.25E+07	3	7.29E+07	33	0.857	0.473
BB	4.23E+07	3	6385335.46	33	6.624	0.001
VT	8.59E+08	3	2.05E+07	33	41.821	0
CMT	1.06E+09	3	3.01E+07	33	35.146	0
TAR	1.34E+09	3	5646281.53	33	236.8	0
OTHERS	1.29E+07	3	2927451.39	33	4.413	0.01

KEY: ***EMB*** *= Earth mud bricks,* ***WB*** *= Wood Baboo,* ***CC*** *= Cement Concrete,* ***BB*** *= Burnt Bricks,* ***VT*** *= Vinyl Tiles,* ***CMT*** *= Ceramic Mable Tiles,* ***TAR*** *=* ***Terrazzo***

Table 15: ANOVA for Clusters on Wall Types

	Cluster		Error		F	Sig.
	Mean Square	Df	Mean Square	Df		
RM	1.93E+12	3	2.53E+10	33	75.959	0
WB	3.72E+09	3	1.38E+09	33	2.7	0.062
STONE	2.26E+08	3	1.63E+08	33	1.381	0.266
CBB	1.32E+12	3	1.66E+10	33	79.306	0
MZ	7.72E+07	3	2.44E+08	33	0.316	0.814
OTHERS	2.62E+07	3	2.02E+07	33	1.294	0.293

KEY: ***RM*** *= Reedmud,* ***WB*** *= Wood Baboo,* ***CBB*** *= Cement Blocks/Bricks,* ***MZ****= Metal Zinc*

5.1 Results on Proximity Matrix

The Proximity matrices on (Table 2 - 4) showed values representing similarity or dissimilarity between each pair of items considering the squared Euclidean distance measure. Looking at Table 2 on roof types, the smallest squared Euclidean distance between roofing tiles and others is 6.051E9 while the smallest Euclidean distance from Table 3 of floor tiles between vinyl tiles and ceramic/Mable tiles is 4.564E8 while the smallest Euclidean distance for wall tiles in Table 4 is stone and other types of wall materials given as 5.507E9.These are the first case to be joined together in each of the building material. The re-computation of the proximity matrix helped in successively identifying the pairs with smallest squared Euclidean distance in each case until optimal pairing is achieved.

This agrees with the unit of Yim and Randeem(2015)who observed that cluster analysis is able to sort out observations or variables into homogeneous groups by reducing several locations or variables with few clases based on materials used in variables with few clusters based on materials used in building constructions.

5.2 Results on Agglomeration Schedules

Agglomerative schedule, Tables 5-7 showed how the variables are clustered together at each stage until all clusters are formed together. The coefficients on the agglomerative schedule indicate the distance between variables/clusters joined together and for a good cluster solution, a jump is observed in the distance coefficient, that is, the stage before the sudden change which indicate the optimal stopping point for merging clusters. From Table 5 it is observed that sudden jump occurs at stage 5. So that there are 8-5=3 clusters formed for roof types. Similar sudden jump is observed on Table 6 for floor types so that the number of clusters formed are 9-5=4cluster. Table 7 shows that stage 2 is where the sudden jump begins as such, the number of clusters formed are 6-2=4. These points of sudden jump are clearly shown on the coefficients of agglomeration schedule.

5.3 Results on Cluster Membership

Table 8 showed cluster memberships with respect to (a) roof types, (b) floor types and (c) wall types of material respectively. From Table 8 the roof type membership showed that Abuja belongs exclusively to cluster1, Lagos to cluster2 while all other 35 states belong to cluster 3. This implies that the type of roof materials used in Abuja and that of Lagos are respectively different from types of roof in other states. Likewise, floor type membership showed that Abuja and 19 states belonged to cluster1 in the use of common floor types of materials. Membership of cluster 2 are 2 states (Kano and Katsina) with common floor types of material used while Lagos exclusively belongs to cluster3, Abia and 13 states belonged to cluster 4.This implies that the type of floor used in cluster 1 differ from the type of floor materials used in cluster 2 , cluster3 and cluster 4. finally, cluster Membership of wall type results showed that Delta state belonged to cluster1, cluster 2 membership includes Rivers and 13 states while cluster 3 membership for wall types of Materials is Abuja and 20 states and cluster 4 belongs to Lagos state.

Similarly, Table 9- 11 on final cluster center showed the type of materials that are predominantly used in these states or locations for construction in Nigeria. The summary of findings is shown in Table 12

.

This result is also consistent with the study by Hepsen and Vatansever (2012) who utilized Hierarchical Agglomerative clustering to develop homogeneous grouping for real estate portfolio. Similarly, it showed the structural relationship among clusters formed just as the work by Abraham *et al*.(1994) who identified similar pattern in US housing market.

5.4 Results of ANOVA

Table 12: Summary of Findings, from Cluster membership and final cluster center

Types of Materials	Clusters Formed for Housing Units in Nigeria	By Their Location	By Types of Materials Used
Roof Types	3clusters Were Formed	Abuja, Lagos and Other 35 Location belongs to cluster (1,2 and 3)	(1 and 3) predominantly use corrugated metal/zinc sheet while cluster 2 predominantly used state asbestos among others.
Floor Types	4 Clusters Were Formed	Abuja and 19 states, Kano and Katsina states, Lagos state, finally Abia and thirteen states belongs to cluster (1,2,3 and 4) respectively	clusters (1,2) Predominantly used earth mud/mudbricks cluster (3,4) used Cement concrete.
Wall Types	4 Clusters Formed	Delta state, Rivers and thirteen (13) states, Abuja and twenty states, Lagos state. belong to Cluster (1,2,3 and 4) respectively	clusters (1 and 3) predominantly used Reed/mud while cluster (2 and 4) used cement block bricks among others

Table 13-15 presents ANOVA on roof types, floor types , and wall types of materials The Analysis of variance (ANOVA) which validated the results in this study showed that cluster classification of the variables provided good classification of roof, floor and wall types of material in Nigeria. By these results, five (5) out of (8) eight roof types, eight (8) out of nine (9) floor types and two (2) out of six (6) wall types are significantly different with p-value < 0.05. This suggests that the cluster classification of the building materials in Nigeria is good. Thus, the use of Agglomerative hierarchical clustering has succeeded in classifying buildings in various locations in Nigeria to units' clusters they belong based on materials used in constructing them. This is almost near the true picture of the scenes in different locations in Nigeria.

6.0 CONCLUSION AND RECOMMENDATION

6.1 Conclusion

From the results of this study, the cluster membership, the agglomerative schedule, and hence final cluster center indicated that: Three (3), Four (4) and Four (4) clusters were formed respectively for roof, floor and wall types of materials. for roof type, the location for cluster1 is Abuja, clusters2 location is Lagos and cluster 3 consists of 35 locations with their associated roof materials for this clusters. Thus, Abuja in cluster 1 and 35 other locations in cluster 3 are found to predominantly used corrugated metal zinc sheet while cluster 2 which is Lagos is found to predominantly used state asbestos.

Likewise, for floor type, the four clusters formed are Abuja and 19 locations as cluster 1, Kano and Katsina as cluster 2, Lagos state as cluster 3 while cluster 4 consists of Rivers and 13 other Locations. The associated floor materials predominantly used for Cluster1 and cluster 2 is earth/mud/bricks while cluster 3 and cluster 4 of floor type is found to predominantly used cement/concrete.

Finally for wall type of materials, the location for the four cluster formed are Delta state as cluster 1, Rivers and 13 other location as cluster 2 , Abuja and 20 location as cluster 3, while Lagos state is in cluster 4, Likewise, the associated wall type of material that is predominantly used in cluster 1 and cluster 3 is Reed/mud among others while cluster 2 and cluster 4 are found to predominantly used Cement/Block/Bricks.

6.2 Recommendations

Based on the findings in this study, the following recommendations are made thus:

(I) The identification of housing materials used in various locations is a clear indication that the available raw materials in such locations can be scientifically improved upon to achieve quality housing.

(ii) The use of corrugated metal zinc sheet and state asbestos for roof, earth mud/mudbricks and cement concrete for floor types, reed/mud and cement/blocks/bricks for wall types, which appears predominantly used in all these locations is a clear indication that building pattern and raw materials differs from one location to another in Nigeria .

(iii) Housing developers are encouraged to use available building materials in their location.

REFERENCES

Abraham, J.M, Goetzmann, W.N & Walter, S.M(1994). Homogeneous groupings of Metropolitan Housing, 3(3): 186-206. https://doi.org/10.1006/jhec.1994.1008

Achunine, B. O. (1992). Urban Relocation: Policy and Practice Evaluation of relocation Programme in Nigeria, Paper presented at the Expert Group Meeting on evaluation of Experience with Relocation Activities, HIS, Rotterdam, the Netherlands, 12-15.

Ahsen, O. and Gulcin P.G. (2005). Space use, dwelling layout and housing quality: an example of low-cost housing in Istanbul: Ashgate publishing Limited, England. Retrieved from Https://www.researchgate.net/publication/27250427.

Allen, M. (2017). Hierarchical Model in the Sage Encyclopedia of Communication Research Methods. DOI: https://dx.doi.org/10.4135/9781483381411.n232: date accessed 15-04-2020

Amole, O.O. (1997). An Evaluation of Students" Residential Facilities in Some Nigerian universities. Unpublished PhD Thesis, submitted to the Department of Architecture Obafemi Awolowo University, Ile-Ife.

Aroni S. & Hasson, S. (1982). "The Meaning of Public Housing in Los Angeles: Contradictory Functions and Values", *ITCC Review.* (44): 21-31.

Bello, A. (2019) Review of the Housing Policies and Programmes in Nigeria. *International Journal of contemporary Research and Review*, 10(2):20603-20616.https://doi.org/10.15520/ijcrr.v10i02.671 Date:20/05/2021

Gallent, N. Shoard, M. Anderson, J. Oades, R. and Tudor C. (2004). Englands Urban Fringes: Multi-Functionality and Planning. Local. Environment. *The International Journal of Justice and Sustainability*. 9(3):217-233. www.researchgate.net>publication

Grimes, Jr. and Orville. F. (1976). Housing for low-income urban families. Economics and policy in the developing world, IBRD Research Publication. Baltimore: Te Johns Hopkins University Press. Pp 190

Henilane, I. (2016). Housing Concept and Analysis of Housing Classification. Baltic *Journal of Real Estate Economics and Construction Management,* 4(1):168-179 DOI:10.1515/bjreecm-2016-0013

Henilane, I. (2015a). The Evaluation of Housing situation in Latvia, In XVI Turiba University international Conference" Towards Smart, Sustainable and inclusive Europe: Challenges for Future Development,"(pp.93-106).Riga, Latvia. ISSN 1691-6069.

Hepsen, A., and Vatansever, A. (2012), Using Hierarchical Clustering Algorithms For Turkish Residential Market, *International Journal of Economics and Finance* 4(1):138.

Johnson M.P. (2006). Decision Models for Affordable Housing and Sustainable Community Development. Carnegie Mellon University, 5000 Forbes Ave. Pittsburgh, PA 15213-3890, Special issue of *Journal of the American Planning Association.JAPAHhousingpaper05.doc 1-27.*

Landau, S. Ster, I. C (2010), Cluster Analysis Overview. Encyclopedia Article.pdf.http://www.researchgate.net/publication/286854317.DOI:10.1016/B978-0-08-044894-7.01315-4. pp72-83

Liao, M. Yunfeng Li. Kianifard, F. Obi, E. and Arcona, S. (2016) Cluster analysis and its application to healthcare claims Data: A study of end-stage renal disease Patients who Initiated Hemolysis. Doi 10.1186/5/288.016-0238-2

Okupe, L. (2002). Private Sector Initiative in Housing Development in Nigeria-how feasible? Housing Today. *The Journal of the Association of Housing Corporations of Nigeria.* 1(6):21-26

Pan, X. Guo, Y. Liu, Z. Zhang, Z. and Shi, Y. (2019). Application of Cluster Analysis and Discriminant Analysis in quality grading of Jadeite red. The second International conference on Physics, Mathematics and Statistics, Iop Conference series: *Journal of Physics: conference Series* 1324(2019) 012101 Doi:10.1088/1742-6596/1324/1/012101:1-7

Rencher, A.C. (2002). Methods of Multivariate Analysis, 2nd Edition, Canada, John Wiley and Sons, Inc.708pp.

Shah, D. Motiani, S and Patel, V. (2013). Analysis of Clustering algorithms on Movie Scripts. Case study: Movie-classification-using K-mean-and- hierarchical-clustering, CorpusID:2790789 . Google.com/amp/S/docplayer.net/amp/21468595

Stafford, D.C., (1978). The Economics of Housing Policy. Groom Helm Ltd. 2-10 St. John Road London SW11. 49(4):577. The town planning review (22/04/21)

Stigler, G. and Becker, G. (1977). *American Economic Review*, 67(2): 76-90.

Soltani, S. and Modarres, R. (2006). Classification of Spatio-temporal pattern of rainfall in Iran using a hierarchical and divisive cluster Analysis, *Journal of Spatial Hydrology*.6(2):1-12

Yang, W. and Miao, R. (2014). The Application of Cluster Analysis in the Classification of China's Modern Service Industries, *Journal of Service Science and Management*, 7: 377-389

Yim, O. and Ramdeem, K. T. (2015). Hierarchical Cluster Analysis: Comparison of Three Linkage Measures and Application to Psychological Data: *The Quantitative Methods For psychology*, 11(1): 8-21