RUHUNA JOURNAL OF SCIENCE
Vol 11 (1): 83-97, June 2020
eISSN: 2536-8400
DOI: http://doi.org/10.4038/rjs.v11i1.89

Ranking and grouping of the districts of Sri Lanka based on the expenses of households: A multivariate analysis

S. R. Gnanapragasam

Department of Mathematics, The Open University of Sri Lanka, Sri Lanka

Correspondence: *srgna@ou.ac.lk;* https://orcid.org/0000-0003-1411-4853

Received: 22nd March 2019, Revised: 09th October 2019, Accepted: 19th April 2020

Abstract. Sri Lanka has twenty-five districts administrated under nine provinces. The cost of living (CoL) diverges among the districts in Sri Lanka like in many parts of the world. Ranking and grouping based on household expenses is useful in decision making by stakeholders. Principal component analysis and cluster analysis are used for ranking and grouping the districts, respectively, based on the expenses of households in Sri Lanka. Sri Lankans spend more for non-food items than food items, particularly for housing and transport, and therefore non-food items are the most influencing factor to decide the CoL in the country. It is concluded that Colombo district has the highest CoL, followed by Gampaha and Kalutara, whereas Kilinochchi and Mullaitivu districts have the least CoL in Sri Lanka. Districts with moderately high CoL were also identified. These classifications will facilitate investors on decision making and also help people to decide which part of the country will be suitable to settle depending on their income and the CoL in districts. Moreover, this grouping will provide some information to policy makers when planning infrastructure development in the country, and it may also provide a direction to use a new index to measure CoL in Sri Lanka.

Keywords: Cluster analysis, cost of living, principal component analysis.

1 Introduction

Sri Lanka - also known as the pearl of the Indian Ocean is one of the top tourist destinations in the world. Twenty-five second-level administrative divisions called 'districts' are administrated under nine first-level administrative divisions or 'provinces' in Sri Lanka. As per the World Bank Report (2018), Sri Lanka is a lower middle-income country with 21.4 million of population. Further, it reports that, its economy grew at an average 5.8% from 2010 to 2017 and it has made significant progress in its socio-economic indicator with social indicators ranked among the highest in South Asia.

In Sri Lanka, the cost of living (CoL) differs among the districts. The consumer price index (CPI) is the most widely used measure of CoL in many countries. It

expresses the overall cost of the several goods and services bought by a typical consumer into a single index for measuring the general price level. The CPI is an accurate measure of the selected goods that make up the typical basket, but it is not a perfect measure of CoL. When CoL increases, on one hand housing becomes less affordable but, on the other hand, it often reflects positive developments at a higher level.

Generally, Colombo consumer price index (CCPI), wholesale price index (WPI) and gross domestic product deflator (GDPD) are the three major indicators used to measure the changes in the prices in Sri Lanka. According to Kulatunge (2017), CCPI is the official CoL index in Sri Lanka. Furthermore, it says that, CCPI covers a large number of items and is heavily weighted to food items which include 41% of the index. Housing, water, electricity and gas weighted 24% while other fuels and transport items weighted 12% of the items in baskets.

Although producing an ideal CoL index may be impossible, that concept directed and encouraged the researchers and relevant sectors to improve the CPI as mentioned by Abraham (2003). In addition, the study of Korale (2001) highlighted the need for a new index to serve as a general inflation indicator. It argued that a specially designed household expenditure survey is needed for a new CPI that should be undertaken primarily for designing weights for the CPI. Jacobs *et al*. (2014) also has pointed out that CPI is not an ideal measure of changes in CoL. Thus, alternative methods must be tried out to measure CoL in Sri Lanka too. An econometric approach to measure CoL was developed by Jorgenson and Slesnick (1990). Gnanapragasam (2016) has classified the districts of Sri Lanka based on CoL by the data obtained in the report of the household income and expenditure survey year 2012/13. In that study, particularly, classifications were done mainly based on the expenditure of essentials non-food items in Sri Lanka. However, districts grew and changed over time and therefore the classifications need to consider the temporal changes as well. This study considers the data, not only, non-food items, but also food items obtained in the report of the Household Income and Expenditure Survey of 2016.

The rankings are useful tools for decision making by stakeholders. Principal component analysis (PCA) is widely used in various research or statistical sectors worldwide. Manage and Scariano (2013) demonstrated the use of PCA in ranking batsmen and bowlers who played in 2012 Indian Pemier League Cricket Tournament. Furthermore, in the studies of Muzamhindo *et al*. (2017) and Steiner (2006), PCA techniques were used to rank the world universities. Clustering is another common tool widely used in many fields of studies such as socio-economic measures (income, education, profession etc.), psychographic measures (interest, life style, motivation, etc.) and measures linked to the buying behavior (price range, buying intensity, buyer, etc.) to identify subgroups within the larger population who share similar pattern on a set of variables. Cluster analysis (CA) is used as a multivariate technique in statistics. The classification could facilitate investors to make their decisions on where to invest to gain more profit while satisfying the needs of customers. Further, this will help people to which part of the country is suitable to settle based on the CoL. Moreover, such grouping will provide some of the required information to policy makers when planning the infrastructure development in the country.

The main aim of this study is to classify the 25 districts based on the expenses on major essential items in the households of Sri Lanka, through two objectives: firstly, to order the districts based on total CoL of a household in Sri Lanka using Principal Component Analysis (PCA), and secondly, to group the districts based on the expenses of a household in Sri Lanka on essential items using the Cluster Analysis (CA).

2 Materials and Methods

2.1 Data and variables

The data for this study have been extracted from the most recent survey report of Household Income and Expenditure released by the Department of Census and Statistics of the Ministry of National Policies and Economic Affairs, Sri Lanka in January 2018. It contains data for all 25 districts in Sri Lanka considering two major categories as food items and non-food items in the survey. In food items category, 15 items have been considered whereas 13 items are taken into account under non-food category. Altogether data for 28 items, as variables, in the households in all 25 districts of Sri Lanka, as observations, are considered. In this study, the food items are taken as X variables and non-food items are taken as Y variables to handle the statistical software conveniently (Table 1).

Table 1. List of variables used in estimating cost of living in districts of Sri Lanka.

Food items		Non- food items	
Variable	*Name*	*Variable*	*Name*
X1	Cereals	**Y1**	Housing
X2	Prepared food	**Y2**	Fuel & Light
X3	Pulses	**Y3**	Personal care & Health expenses
X4	Vegetables	**Y4**	Transport
X5	Meat	**Y5**	Communication
X6	Fish	**Y6**	Education
X7	Dried fish	**Y7**	Cultural activities and entertainments
X8	Eggs	**Y8**	Household non-durable goods & services
X9	Coconuts	**Y9**	Clothing textiles & Footwear
X10	Condiments	**Y10**	Household durable goods
X11	Milk & milk food	**Y11**	Other miscellaneous expenses
X12	Fats & oil	**Y12**	Other adhoc (rarely) expenses
X13	Sugar, Jaggery & Treacle	**Y13**	Liquor, Drugs & Tobacco
X14	Fruits		
X15	Other food items		

2.2 Data analysis

Descriptive statistics

The descriptive statistics such as mean, standard deviation, minimum, maximum and total of all the items are summarized. District-wise average expenditures on food and non-food items are compared and, average expenditures on all 28 items for each district are reported.

Correlation analysis

The associations among the variables on food and non-food items, and the relationship between these items in both categories in terms of correlations were analysed. The p-value of Pearson's correlation is obtained to observe the strength of the relationship among those 28 items, 5% level of significant is considered.

Principle Component Analysis (PCA)

Using the principal component as an index requires the determination of principal component scores. The principal component scores were obtained by substituting the observed values of the standardized variables into the following equation of the ith principal component (PC_i).

$$PC_i = e_i' Z \quad (1)$$

where, e_i'=Transpose of i^{th}eigenvector and Z= Matrix form of vectors of standardized variables. In PCA, the first few components (usually the first and the second or third) explain the greater percentage of the variance of original data. Therefore, in this study, only the first principle component score is used to rank the districts based on the total CoL.

Cluster Analysis (CA)

The cluster analysis is used to classify similar districts into homogeneous clusters. To determine the number of clusters, in this study, firstly the rule of thumb "square root of the half of the total number of observations" is chosen for convenient. The classification is based on a particular distance measure (such as Euclidean distance) between the clusters in terms of similarity among them. Secondly, in this study, to implement CA, the distance between clusters is calculated by consideration of Squared Euclidian distance criterion. To choose the appropriate linkage method is also crucial in CA. In order to identify the groups, thirdly, the following linkage methods are taken into consideration in this study: Single linkage, Average linkage, Complete linkage, Centroid linkage and Ward's linkage. The appropriate cluster is selected based on the criteria made by the majority of five linkage methods listed above. Priority was given to complete, centroid and average linkages as clusters do not have equal sizes.

3 Results and Discussion

3.1 Average expenditure

The descriptive statistics of the expenses of the items (Table 2) and the district wise average expenditures on all the categories (Figure 1) are reported.

Table 2: Descriptive statistics of the expenses of the items in Sri Lanka.

Item	Expenses in LKR				
	Average	Std Dev	Min	Max	Total
Cereals	3,109.00	501.00	2,385	4,527	77,731
Prepared food	1,928.00	783.00	980	4,394	48,202
Pulses	668.80	123.10	363	922	16,719
Vegetables	1,788.40	259.10	1,078	2,101	44,709
Meat	944.00	437.80	418	1,982	23,601
Fish	1,959.00	768.00	909	3,772	48,980
Dried fish	614.00	313.10	49	1,084	15,350
Eggs	204.40	31.65	156	284	5,110
Coconuts	1,053.00	169.60	701	1,469	26,326
Condiments	1,836.50	310.50	1,303	2,486	45,913
Milk & milk food	1,415.00	351.40	932	2,668	35,374
Fats & oil	494.80	109.90	321	691	12,371
Sugar, Jaggery & Treacle	462.70	82.10	346	662	11,567
Fruits	541.50	188.80	179	1,226	13,538
Other food items	1,411.10	376.30	922	2,626	35,278
Housing	4,964.00	3,513.00	1,841	19,232	124,097
Fuel & Light	1,574.00	508.00	928	3,419	39,355
Personal care & Health expenses	2,121.00	929.00	673	4,782	53,015
Transport	3,648.00	1,567.00	1,156	9,483	91,190
Communication	927.00	320.00	558	2,072	23,175
Education	1,706.00	711.00	820	4,169	42,644
Cultural activities and entertainments	681.70	443.90	66	1,806	17,043
Household non-durable goods & Household services	528.60	238.40	292	1,518	13,214
Clothing textiles & Foot wear	1,506.90	301.90	738	2,061	37,673
Household durable goods	2,002.00	1,130.00	440	4,286	50,039
Other miscellaneous expenses	5,115.00	2,038.00	1,809	8,591	127,870
Other adhoc (rarely) expenses	4,277.00	1,816.00	1,292	8,358	106,917
Liquor, Drugs & Tobacco	999.20	377.10	310	2,148	24,980

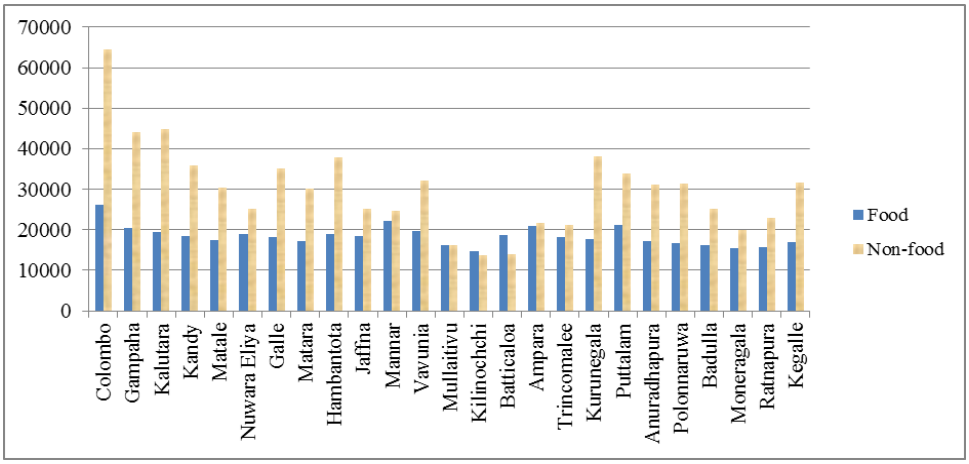

Fig. 1. District wise average expenditures on food and non-food items

It is highlighted from Figure 1 that in almost all the districts, expenditures on non-food items is higher than on food items. Further, less variation in expenditures on food items between the districts can be seen. In contrast, more variation can be seen in the expenditure on non-food items among the districts. It indicates that, CoL mainly depends on the expenditure on non-food items as expenditure on food items has less variation in Sri Lanka. Moreover, expenditures on both categories (LKR 26,066 for food items and LKR 64,604 for non-food items) are very high in Colombo district whereas which (LKR 14,688 for food items and LKR 13,795 for non-food items) are very low in Kilinochchi district. Therefore, the ranges of expenditures on food items and non-food items between districts are LRK 11,378 and LKR 50,809 respectively.

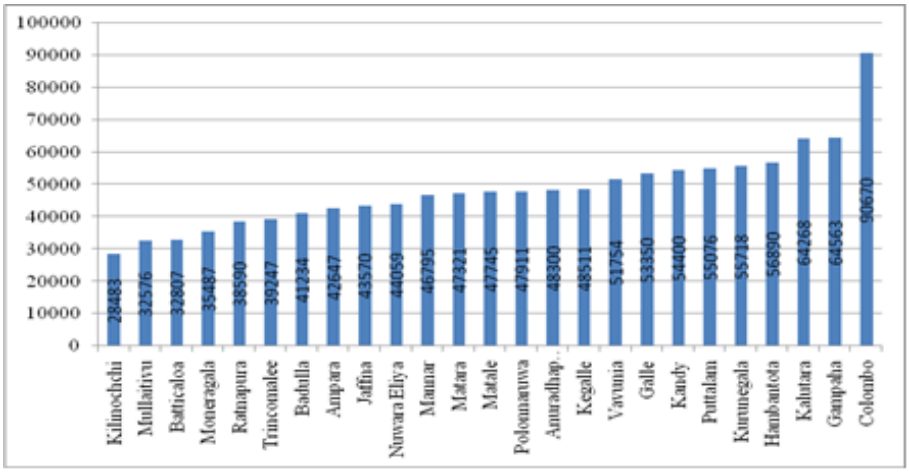

Fig. 2. District wise average expenditures on all items

Figure 2 illustrates district wise average expenditures of a household on both categories. It clearly indicates that total expenditures in Colombo district is much more higher than other districts as the gap between Colombo district (1st rank) and Gampaha district (2nd rank) is LKR 26,107 and it is relatively high with other gaps between consequent (in terms of the ascending order of expenditures) districts. Also, the average expenditure in all three districts of the Western province of Sri Lanka are having top three in the order (this univariate ordering is based on only average expenditure of all the items taken in the survey).

When the monthly average expenditure is considered (Figure 2), the difference between Colombo with other districts in the Western province is over LKR 26,000. Hence, it can be stated by comparing with expenditures in other districts too that, Colombo district has an extreme value in terms of the total expenditure per month. At the same time, expenditure is manageable in the districts of Kilinochchi, Mullaitivu and Batticaloa (they receive the last three ranks in non-food and all items categories) with roughly from LKR 28,500 to LKR 33,000. But approximately triple this amount is needed to survive in district of Colombo and nearly double of this amount is necessary to live in the districts of Gampaha and Kalutara which have the next highest amounts of expenditure. Therefore, it suggests that Colombo district has an extreme value for the monthly expenditure. On the average, LKR 48,500 per month is required to live in Sri Lanka.

3.2 Associations among the items

The relationships, in terms of the p-value of Pearson's correlation, amongst the variables (items) considered in this study are reported for all the categories separately as above (Table 3). It is observed that the pair wise correlation between some of the variables in the category of food items is significant ($p<0.05$). The relationship among the non-food items are summarized in Table 4 in terms of the p-values of Pearson's correlation.

Table 3: Probabilities for correlation among the variables in food items (p values in bold font are significant).

p-value	X1	X2	X3	X4	X5
X1	1.000				
X2	**0.031**	1.000			
X3	0.075	0.527	1.000		
X4	0.367	0.403	0.147	1.000	
X5	0.512	**0.021**	**0.007**	0.363	1.000
X6	0.908	0.057	**0.001**	**0.008**	**0.000**
X7	**0.022**	0.108	0.156	**0.001**	0.212
X8	0.128	0.944	0.880	0.074	**0.012**
X9	0.191	0.141	0.146	0.083	0.136
X10	0.903	**0.004**	0.836	0.166	**0.016**
X11	0.582	**0.000**	0.072	0.113	0.215
X12	**0.000**	0.532	0.493	0.466	0.066
X13	0.082	0.792	0.253	**0.004**	**0.007**

X14	0.127	**0.000**	0.328	**0.025**	0.309
X15	0.558	**0.001**	0.541	0.149	**0.025**

p-value	X6	X7	X8	X9	X10
X6	1.000				
X7	**0.006**	1.000			
X8	0.595	0.962	1.000		
X9	0.502	**0.036**	0.301	1.000	
X10	**0.021**	0.807	**0.046**	0.132	1.000
X11	0.687	0.168	**0.039**	0.384	**0.006**
X12	0.834	**0.029**	0.116	**0.001**	0.860
X13	**0.000**	**0.000**	0.545	0.324	0.128
X14	0.820	0.100	0.428	0.166	**0.004**
X15	0.480	0.199	**0.046**	0.496	**0.001**

p-value	X11	X12	X13	X14	X15
X11	1.000				
X12	0.570	1.000			
X13	0.854	0.132	1.000		
X14	**0.000**	0.800	0.854	1.000	
X15	**0.000**	0.350	0.759	**0.000**	1.000

Table 4: Probabilities for correlation among the variables in non-food items (p values in bold font are significant).

p-value	Y1	Y2	Y3	Y4	Y5
Y1	1.000				
Y2	**0.000**	1.000			
Y3	**0.000**	**0.000**	1.000		
Y4	**0.000**	**0.000**	**0.000**	1.000	
Y5	**0.000**	**0.000**	**0.000**	**0.000**	1.000
Y6	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
Y7	**0.000**	**0.017**	**0.000**	**0.000**	**0.000**
Y8	**0.000**	**0.000**	**0.001**	**0.000**	**0.000**
Y9	**0.031**	**0.007**	**0.008**	**0.003**	**0.006**
Y10	0.199	0.469	0.122	**0.046**	0.161
Y11	**0.008**	0.141	**0.001**	**0.000**	**0.004**
Y12	**0.051**	0.254	**0.007**	**0.004**	**0.033**
Y13	0.333	0.861	0.205	0.062	0.485

p-value	Y6	Y7	Y8	Y9	Y10
Y6	1.000				
Y7	**0.000**	1.000			
Y8	**0.000**	**0.001**	1.000		
Y9	**0.028**	**0.020**	**0.027**	1.000	
Y10	**0.059**	0.074	0.202	0.256	1.000
Y11	**0.002**	**0.001**	**0.026**	**0.026**	**0.020**
Y12	**0.014**	**0.004**	0.107	0.069	**0.000**
Y13	0.320	0.065	0.260	0.069	0.240

p-value	Y11	Y12	Y13
Y11	1.000		
Y12	**0.000**	1.000	
Y13	**0.038**	**0.037**	1.000

From Table 4 it concludes with 95% confidence that most of the variables among non-food items are highly correlated ($p < 0.05$). Thus, it suggests carrying out the multivariate analysis.

It is important to look at the relationship between variable in the categories of food items versus non-food items. Thus, Table 5 provides the p-values of Pearson's correlation between the variables of cross categories food items and non-food items.

Table 5. Correlation between the variables in food (X) and non-food (Y) items (p values in bold font are significant).

p-value	X1	X2	X3	X4	X5
Y1	0.130	**0.000**	0.171	**0.020**	0.618
Y2	0.689	**0.000**	0.781	0.678	**0.004**
Y3	0.546	**0.000**	**0.047**	**0.055**	0.858
Y4	0.275	**0.000**	0.092	**0.023**	0.770
Y5	0.732	**0.000**	0.126	0.274	0.175
Y6	0.339	**0.000**	**0.033**	0.218	0.754
Y7	0.214	**0.005**	**0.024**	**0.005**	0.356
Y8	0.385	**0.000**	0.169	0.156	0.462
Y9	0.595	**0.033**	0.496	**0.019**	**0.019**
Y10	**0.019**	0.104	0.359	0.825	0.514
Y11	0.450	0.073	**0.029**	**0.025**	0.257
Y12	0.242	0.061	0.142	0.233	0.369
Y13	0.719	0.856	**0.030**	**0.002**	0.082
p-value	X6	X7	X8	X9	X10
Y1	0.804	**0.010**	0.405	0.093	**0.047**
Y2	**0.048**	0.679	0.193	0.480	**0.003**
Y3	0.822	0.077	0.974	**0.006**	**0.006**
Y4	0.565	**0.007**	0.753	**0.026**	**0.035**
Y5	0.559	0.099	0.451	0.150	**0.032**
Y6	0.728	**0.049**	0.733	0.089	0.127
Y7	0.204	**0.002**	0.865	**0.001**	0.161
Y8	0.827	0.107	0.567	0.254	0.097
Y9	0.533	0.198	**0.010**	0.288	**0.002**
Y10	0.773	**0.009**	0.295	0.093	0.308
Y11	0.147	**0.000**	0.725	**0.000**	0.521
Y12	0.687	**0.026**	0.389	**0.001**	0.231
Y13	**0.019**	0.121	0.919	0.478	0.348
p-value	X11	X12	X13	X14	X15
Y1	**0.000**	0.850	**0.051**	**0.000**	**0.000**
Y2	**0.000**	0.258	0.481	**0.000**	**0.000**
Y3	**0.000**	0.671	0.662	**0.000**	**0.001**
Y4	**0.000**	0.981	0.156	**0.000**	**0.000**
Y5	**0.000**	0.526	0.643	**0.000**	**0.000**
Y6	**0.000**	0.849	0.408	**0.000**	**0.000**
Y7	**0.001**	0.265	0.084	**0.000**	**0.018**
Y8	**0.000**	0.315	0.244	**0.000**	**0.000**
Y9	**0.002**	0.351	0.266	**0.001**	**0.000**
Y10	0.169	**0.042**	0.893	**0.025**	**0.035**
Y11	**0.042**	0.114	0.121	**0.020**	**0.049**
Y12	0.131	0.074	0.755	**0.014**	0.087
Y13	0.431	0.353	0.222	**0.056**	0.308

3.3. Ranking the districts of Sri Lanka using first principal component

Since some of the previous studies (Salmond and Crampton 2002, Houweling *et al*. 2003, McKenzie 2005, Messer *et al*. 2006, Primpas *et al*. 2010) suggested that only the first principal component is indeed providing a better measure of their relevant studies, only the first principle component score is used to rank the districts based on the total CoL in this study. Based on the first principal component scores, the districts are ranked in three different categories such as ranks on food items, ranks on non-food items and ranks on all items separately. According to the ranks appear in Table 6, it is clearly noted that, Colombo district has the 1st rank whereas Mullaitivu and Kilinochchi have the last two 24th and 25th ranks, respectively, in all the categories. In addition, ranks assigned to Trincomalee district are same in all three categories as 21st.

Table 6: Ranks of districts in Sri Lanka based on cost of living.

Province	District	Rank		
		Food Items	Non-Food Items	All Items
Western	*Colombo*	1	1	1
	Gampaha	5	2	2
	Kalutara	4	3	3
Central	*Kandy*	8	4	4
	Matale	13	12	11
	Nuwara Eliya	11	18	15
Southern	*Galle*	7	8	7
	Matara	12	15	13
	Hambantota	6	6	5
Northern	*Jaffna*	22	14	18
	Mannar	16	16	17
	Vavunia	9	7	9
	Mullaitivu	24	24	24
	Kilinochchi	25	25	25
Eastern	*Batticaloa*	14	23	23
	Ampara	3	20	14
	Trincomalee	21	21	21
North Western	*Kurunegala*	10	5	8
	Puttalam	2	9	6
North Central	*Anuradhapura*	17	11	12
	Polonnaruwa	20	13	16
Uwa	*Badulla*	19	17	19
	Moneragala	23	22	22
Sabaragamuwa	*Ratnapura*	18	19	20
	Kegalle	15	10	10

Further, it is noted that, the ranks in the categories of non-food items and all items are similar to each other compared with the ranks in the other category of food items. Thus, it also indicates that the non-food items influence more on the total expenditure of a

household in Sri Lanka. Ampara district rank is 3rd in food items category (within top three), however, for non-food items category, it receives 20th rank. Moreover, Hambantota and Mannar districts receive the same ranks in both categories of food items and non-food items. Also, they receive very closer rank in all items category such as 6th to 5th and 16th to 17th.

When only the category of all items is concerned in Table 6, Kandy has the highest rank as 4th among the three districts in the Central province while Hambantota, Puttalam, Vavunia, and Kegalle lead the other districts in Southern, North Western, Northern and Sabaragamuwa provinces respectively; they fall in top 10 ranks. Both districts in Uwa province (Badulla and Moneragala) have relatively lower ranks. Except Vavuniya in Northern province, other 4 districts have moderately lower ranks. It is further observed from Table 6 that districts in the Western province receive top three ranks in category of non-food items as well as in category of all items. Here it is noted that, the first principal component score is also very high for Colombo district (-12) and the next scores of Gampaha and Kalutara are -4.03 and -3.26 respectively. Therefore, it can be concluded that Colombo is the mostly expensive district followed by Gampaha and Kalutara whereas Kilinochchi and Mullaitivu are the least expensive districts in Sri Lanka.

3.4 Grouping the districts of Sri Lanka using cluster analysis

The importance of the clustering in the multivariate analysis are discussed in previous studies (Anderberg 1973, Kaufmann and Rousseeuw 1990, Peneder 2005). Also, the prior specifications about the number of groups are not in CA but mostly the appropriate number of groups is determined within the clustering method. Determining the number of clusters is crucial and also subjective in most of the time. It is noted that the total number of observations in this study is 25 as the data is available for 28 variables in all 25 districts of Sri Lanka. The number of clusters $\sqrt{n/2}$ is 3.54 (n is the number of observations (districts)) for this data. Therefore 3 or 4 clusters can be considered as it falls in between those two integers. Gnanapragasam (2017) observed that there were no changes in first two clusters when 3-clusters were needed and only the 3rd and 4th cluster had to be merged from 4-cluster groups. Also, to merge the last 2 clusters, only the district Kilinochchi (singleton in 4th group) must be joined with other 21 districts. Accordingly, in this study, only 4-cluster groupings are considered. When comparing the combinations of five linkage methods listed, previous studies have concluded that there is no one superior method in all situations but it depends on the form of the data (Kuiper and Fisher 1975, Blashfield 1976, Jain *et al*. 1986, Hands and Everitt, 1987, Johnson and Wichern 2002, Ferreira and Hitchcock 2009). Also, Ward's and complete linkages worked best for clusters of equal sizes, but for unequal cluster sizes, centroid and average linkages worked the best. In the present study, three categories are considered as food items, non-food items and all items separately. The five linkage methods, namely, Average linkage, Centroid linkage, Complete linkage, Single linkage and Ward's linkage with squired Euclidean distance measure are considered. The suitable cluster is selected based on the suggestions made by the

majority of 5 linkage methods. Accordingly, the Table 7, Table 8 and Table 9 are summarized for three categories separately. Table 7 shows that Colombo, Nuwara Eliya and Mannar districts individually having separate clusters and all other 22 districts fall into a single cluster when they group into 4-clusters based on the expenses on food items. It also indicates that, expenses on food items have less variation among most of the districts of Sri Lanka.

Table 7: Four-cluster groupings of districts for the category of food items.

Cluster 1 (Most expensive district)	Colombo			
Cluster 2 (High level Expensive districts)	Gampaha Kalutara Kandy Matale Galle Matara	Hambantota Jaffna Vavunia Mullaitivu Kilinochchi Batticaloa	Ampara Trincomalee Kurunegala Puttalam Anuradhapura	Polonnaruwa Badulla Moneragala Ratnapura Kegalle
Cluster 3 (Moderate level Expensive districts)	Nuwara Eliya			
Cluster 4 (Least expensive Districts)	*Mannar*			

When the expenses on non-food items are considered, it is observed from Table 8 that, Colombo district again belongs to a separate cluster. Gampaha, Kalutara and Kandy districts are having a separate cluster whereas all other 21 districts split into 2 different clusters. Also, it is seen from Table 8 that, Nuwara Eliya and Mannar district falls into the same cluster (based on the non-food items) unlike in Table 7 (based on the food items).

Table 8: Four-cluster grouping of districts for the category of non-food items.

Cluster 1 (Most expensive District)	*Colombo*		
Cluster 2 (High level Expensive districts)	*Gampaha*	*Kalutara*	*Kandy*
Cluster 3 (Moderate level Expensive districts)	*Matale* *Galle* *Matara* *Hambantota*	*Vavunia* *Kurunegala* *Puttalam* *Anuradhapura*	*Polonnaruwa* *Kegalle*
Cluster 4 (Least expensive Districts)	*Nuwara Eliya* *Jaffna* *Mannar* *Mullaitivu*	*Kilinochchi* *Batticaloa* *Ampara* *Trincomalee*	*Badulla* *Moneragala* *Ratnapura*

Table 9: Four-cluster groupings of districts for the category of all items.

Cluster 1 (Most expensive District)	*Colombo*	
Cluster 2 (High level Expensive districts)	*Gampaha*	*Kalutara*
Cluster 3 (Moderate level Expensive districts)	*Kandy* *Matale* *Galle* *Matara* *Hambantota* *Vavunia*	*Kurunegala* *Puttalam* *Anuradhapura* *Polonnaruwa* *Kegalle*
Cluster 4 (Least expensive Districts)	*Nuwara Eliya* *Jaffna* *Mannar* *Mullaitivu* *Kilinochchi* *Batticaloa*	*Ampara* *Trincomalee* *Badulla* *Moneragala* *Ratnapura*

According to the groupings appear in Table 9 based on the expenditure on all items (food and non-food items), here too Colombo district is having a separate cluster. The 2nd cluster now has only Gampaha and Kalutara districts whereas Kandy district joins with the districts in 3rd cluster in Table 8. Further, it is noted that the districts in cluster 4 is the same in both Table 8 and Table 9. Thus, the change of clusters between non-food items and all items is only the Kandy district is shifted from cluster 2 in Table 8 to cluster 3 in Table 9. It can be observed a similarity in groupings, except Kandy district, among the clusters in both non-food items and all items categories.

In the clusters from all 3 tables, Colombo district is isolated from all other districts. Also, based on the ranking the districts of Sri Lanka (section 3.3), Colombo district had exceptional value for total expenses (PC score is also relatively high for Colombo district). Thus, Colombo district is considered here as an extreme observation, and therefore, the groupings of clustering was re-processed by removing Colombo district data from the original data set. The purpose of this process is to check whether there are any differences in groupings of those 24 districts in Sri Lanka. For this purpose, only 3-cluster groupings are obtained, (results not reported) and clearly indicated the same groupings for the districts, except Colombo, like in Table 7, Table 8 and Table 9 in all the categories. Therefore, it can be confirmed that the 4-cluster groupings among the districts irrespective of Colombo district is inclusive in the data.

4 Conclusions

It is concluded that, on average, Sri Lankans spend more for the non-food items than spending for the food items, particularly, for the housing and then for transport. Colombo district is the mostly expensive district followed by Gampaha and Kalutara whereas Kilinochchi and Mullaitivu are the least expensive districts in Sri Lanka. Since

it is observed that almost similar groupings appear from the non-food items and all-items categories, it can be concluded that, non-food items mostly influence the CoL in Sri Lanka. These classifications will facilitate investors to make their decision on where to invest to gain more profit while satisfying the need of customers in that district. Further, this will help the people to decide when settling which part of the country will suit to decrease their CoL. Moreover, this grouping will provide some of the required information to policy makers when planning the infra-structure development in the country and it also may provide a direction to a new index to measure CoL in Sri Lanka.

Acknowledgments

Two anonymous reviewers are acknowledged for valuable comments on the initial draft of the manuscript.

Supporting Material:

1. **Datasets used in the analysis** can be extracted from: http://repo.statistics.gov.lk/bitstream/handle/1/784/HIES2016_FinalReport.pdf?sequence=1&isAllowed=y (In page numbers 24 and 28 in the final report of Household Income and Expenditure for food items and non-food items respectively- ISBN 978-955-702-054-9)
2. ***MINITAB* statistics package is used to PCA and CA and the projects can be seen from:** https://drive.google.com/drive/folders/1sACJ68J6XBgh2hJWsB5cDgETHTX529zg

References

Abraham KG. 2003. Towards a cost of living index: Progress and prospects. *Journal of Economic Perspectives* 17(1): 45-58.

Anderberg MR. 1973. *Cluster analysis for applications*, Academic Press, New York.

Blashfield RK. 1976. Mixture model tests of cluster analysis: Accuracy of four agglomerative hierarchical methods. *The Psychological Bulletin* 83(3): 377–388.

Ferreira L, Hitchcock DB. 2009. A comparison of hierarchical methods for clustering functional data. *Journal of Communications in statistics- Simulation and computation* 38(9):1925–1949.

Gnanapragasam SR. 2016. Classification of districts in Sri Lanka based on the cost of living: A multivariate approach. *Proceedings of the Wayamba University International Conference, Sri Lanka 2016*: 18.

Gnanapragasam SR. 2017. A multivariate approach to classify the districts of Sri Lanka based on the cost of living. *International Journal of Information Research and Review* 4(5): 4128-4132.

Hands S, Everitt B. 1987. A Monte Carlo study of the recovery of cluster structure in binary data by hierarchical clustering techniques. *Multivariate Behavioral Research*. 22(2): 235–243. DOI:10.1207/s15327906mbr2202_6.

Houweling TAJ, Kunst AE, Mackenbach JP. 2003. Measuring health inequality among children in developing countries: does the choice of the indicator of economic status matter? *International Journal for Equity in Health* 2(8)

Jacobs D, Perera D, Williams T. 2014. Inflation and the cost of living. *RBA Bulletin, Reserve bank of Australia*, March Quarter 2014: 33-46.

Jain NC, Indrayan A, Goel LR. 1986. Monte Carlo comparison of six hierarchical clustering methods on random data. *Pattern Recognition* 19 (1): 95-99.

Johnson RA, Wichern DW. 2002. *Applied multivariate statistical analysis*. Upper Saddle River, NJ: Prentice Hall.

Jorgenson DW, Slesnick DT. 1990. Individual and social cost of living indexes. *Contribution to Economic Analysis* 196: 155-234. DOI: 10.1016/B978-0-444-88108-3.50009-3

Kaufman L, Rousseeuw PJ. 1990. *Finding groups in data. An introduction to cluster analysis*, Wiley, New Jersey.

Korale RBM. 2001. The problem of measuring cost of living in Sri Lanka, Research Studies. *Macroeconomic Policy and Planing Series, IPS Publication*. Retrieved from http://www.ips.lk/wp-content/uploads/2017/01/Problems-of-Measuring-Cost-of-Living.pdf

Kuiper FK, Fisher L. 1975. A Monte Carlo comparison of six clustering procedures. *International Biometric Society* 31(3): 777–783. DOI: 10.2307/2529565

Kulatunge S. 2017. Inflation dynamics in Sri Lanka: An empirical analysis, *Central bank of Sri Lanka – Staff studies* 45 (1 & 2): 31-66.

Manage ABW, Scariano SM. 2013. An introductory application of principal components to cricket data, *Journal of Statistics Education*. 21(3): DOI:10.1080/10691898.2013.11889689

McKenzie DJ. 2005. Measuring inequality with asset indicators. *Journal of population economics* 18(2): 229-260.

Messer LC, Laraia BA, Kaufman JS, Eyster J, Holzman C, Culhane J, Elo I, Burke JG, O'Campo P. 2006. The development of a standardized neighborhood deprivation index. *Journal of Urban Health* 83(6): 1041-1062. DOI:10.1007/s11524-006-9094-x

Muzamhindo S, Kong Y, Famba T. 2017. Principal component analysis as a ranking tool - A case of world universities. *International Journal of Advanced Research* 5(6): 2114-2135. DOI :10.21474/IJAR01/4650

Peneder M. 2005. Creating industry classifications by statistical cluster analysis, *Estudios De Economia Aplicada* 23(2): 451-463.

Primpas I, Tsirtsis G, Karydis M, Kokkoris GD. 2010. Principal component analysis: Development of a multivariate index for assessing eutrophication according to the European water framework directive. *Ecol Indic* 10(2): 178-83. DOI: 10.1016/j.ecolind.2009.04.007.

Salmond C, Crampton P. 2002. NZDep2001 index of deprivation. Department of public health, Wellington School of Medicine and Health Sciences. Retrieved from: https://www.researchgate.net/publication/228798047

Steiner JE. 2007. World university rankings: A principal component analysis. *Instituto de Estudos Avançados Instituto de Astronomia, Geofísica e Ciências Atmosféricas Universidade de São Paulo*. Retrieved from http://arxiv.org/ ftp/physics/papers/ 0605/0605252.pdf.

World Bank in Sri Lanka. 2018. Retrieved from https://www.worldbank.org/en/country/ srilanka/overview (accessed 22-03-2019).